ESTIMASI PARAMETER DISTRIBUSI WEIBULL TERSENSOR-II MENGGUNAKAN MLE

skripsi

disajikan sebagai salah satu syarat

untuk memperoleh gelar Sarjana Sains

Program Studi Matematika

oleh

Inang Amiril Mukminin

4150408015

JURUSAN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS NEGERI SEMARANG

2015

PERNYATAAN

Saya menyatakan bahwa skripsi ini bebas plagiat, dan apabila di kemudian hari terbukti terdapat plagiat dalam skripsi ini, maka saya akan bersedia menerima sanksi sesuai ketentuan perundang-undangan.

Semarang, Agustus 2015

METERAI TEMPEL 3DE13ADF392152544 6000 ENAM RIBU RUPIAH

Inang Amiril Mukmini
NIM 4150408015

PENGESAHAN

Skripsi yang berjudul

Estimasi Parameter Distribusi Weibull Tersensor-II Menggunakan MLE.

disusun oleh

Inang Amiril Mukminin

4150408015

telah dipertahankan di hadapan sidang Panitia Ujian Skripsi FMIPA UNNES pada tanggal 27 Agustus 2015.

Panitia:

Ketua

Prof. Dr. Wiyanto, M. Si.
19631012198803 1001

Sekretaris

Drs. Arief Agoestanto, M. Si.
196807221993031005

Ketua Penguji

Putriaji Hendikawati, S.Si., M.Pd., M.Sc.
198208182006042001

Anggota Penguji/
Pembimbing Utama

Dr. Scolastika Mariani, M.Si.
196502101991022001

Anggota Penguji/
Pembimbing Pendamping

Drs. Sugiman, M.Si.
196401111989011001

MOTTO DAN PERSEMBAHAN

Motto:

"Life is funs and games until someone frozen solid".

Persembahan:

Orang tua terhebat dalam hidupku Surati dan Rubai, terima kasih atas segalanya.

Kakak-kakakku Rozikin, Khotimah dan Rojanah serta adikku Uswatun, kalian saudara terbaik yang Allah SWT kirimkan untukku, terima kasih seluruh dukungannya.

Topan, Nicko, sahabat terbaik yang Allah SWT perkenalkan padaku.

Ainotenshi, yang telah memberiku arah untuk tetap berjuang dan bertahan dalam kehidupan.

Anak-anak kos Tazkya, teman-teman seperjuangan.

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Tuhan Yang Maha Esa yang telah melimpahkan berkah serta hidayah-Nya kepada penulis sehingga skripsi ini dapat terselesaikan. Sholawat serta salam semoga senantiasa tercurah kepada suri teladan yang mulia, Nabi Muhammad SAW yang telah memberikan tuntunan yang bijaksana untuk umat manusia umumnya dan pada penulis khususnya.

Terselesaikannya skripsi ini, merupakan sebuah usaha dan perjuangan yang berlandaskan keteguhan, kesabaran, dan keikhlasan. Terima kasih atas kemurahan dari kekuasaan-Nya yang tidak tertandingi oleh apapun dan siapapun.

Penyusunan skripsi ini tidak terlepas dari berbagai pihak yang dari awal hingga akhir memberikan segenap dukungan, baik moral dan spiritual. Hanya ucapan terima kasih yang bisa penulis haturkan kepada pihak-pihak yang selalu memberikan dukungan tenaga, pikiran, dan semangat.

1. Prof. Dr. Fathur Rokhman, M.Hum., Rektor Universitas Negeri Semarang.
2. Prof. Dr. Wiyanto, M.Si., Dekan Fakultas MIPA Universitas Negeri Semarang.
3. Drs. Arief Agoestanto, M.Si., Ketua Jurusan Matematika Fakultas MIPA Universitas Negeri Semarang.
4. Dra. Kristina Wijayanti, M.S., Ketua Program Studi Matematika Fakultas MIPA Universitas Negeri Semarang.
5. Dr. Scolastika Mariani, M.Si serta Drs. Sugiman, M.Si, dosen pembimbing yang telah mencurahkan segenap bimbingan, kesabaran, dan pengertian

kepada penulis dari awal penyusunan sampai akhir selesainya skripsi ini. Mohon maaf jika selama ini banyak sikap yang kurang berkenan di hati Ibu dan Bapak.

6. Putriaji Hendikawati, S.Si., M.Pd., M.Sc. Terimakasih atas inspirasi dan semangat yang telah Ibu bagikan kepada penulis, sehingga semua ini bisa tercapai.
7. Seluruh Dosen di Fakultas MIPA Universitas Negeri Semarang yang telah membagikan banyak ilmu tentang berbagai hal kepada penulis.
8. Bapakku Rubai dan Ibuku Surati yang selalu memberikan kekuatan dan inspirasi untuk tetap berjuang.

Berbagai saran maupun kritik demi penyempurna lebih lanjut atas penelitian pengembangan skripsi ini sangat diharapkan oleh penulis. Semoga memberi manfaat bagi penulis dan bagi pembaca.

Semarang,   Agustus 2015

Penulis

ABSTRAK

Amirilmukminin, Inang. 2015. *Estimasi Parameter Distribusi Weibull Tersensor-II Menggunakan MLE*. Skripsi, Jurusan Matematika FMIPA UNNES. Pembimbing utama Dr. Scolastika Mariani, M.Si., dan Pembimbing Pendamping Drs. Sugiman, M.Si.

Kata kunci : estimasi, distribusi weibull, mle

Inferensi statistik adalah proses yang menggunakan sampel statistik untuk menduga atau menaksir hubungan parameter populasi. Pengambilan kesimpulan tentang parameter-parameter populasi berdasarkan data sampel yang diambil dari populasi. Dengan demikian penaksiran parameter merupakan suatu metode yang digunakan untuk memprediksi karakteristik suatu populasi berdasarkan sampel yang diambil.

Salah satu distribusi yang sering digunakan dalam uji data hidup adalah distribusi Weibull yang diperkenalkan oleh fisikawan Swedia yaitu Waloddo Weibull. Distribusi Weibull selama bertahun-tahun menjadi salah satu model data statistik yang memiliki jangkauan luas dalam uji data hidup dan dengan kelebihan utama dalam keakuratan kegagalan.

Salah satu metode estimasi adalah metode estimasi maksimum likelihood. Maksimum likelihood mendasarkan inferensinya pada sampel, dan juga metode ini salah satu metode untuk menaksir parameter berdistribusi Weibull. Ide dasar metode estimasi maksimum likelihood adalah mencari nilai parameter yang memberi kemungkinan (likelihood) paling besar untuk mendapatkan data observasi sebagai estimator dan digunakan untuk menentukan parameter yang memaksimalkan kemungkinan dari data sampel. Metode maksimum likelihood, metode estimasi parameter lebih mudah, sehingga banyak digunakan. Akan tetapi metode ini hanya dapat digunakan bilamana distribusi populasi diketahui. Selain itu, metode maksimum likelihood sangat sensitive terhadap data ekstrim sehingga berpengaruh terhadap nilai rata-rata maupun variansi.

Pengamatan berpusat pada bagaimana menentukan estimator parameter α dan β pada distribusi Weibull tersensor tipe II. Metode estimasi yang dipakai adalah metode estimasi maksimum likelihood, melalui estimasi data historik tabel mortalitas CSO 1941. Menggunakan estimasi maksimum likelihood untuk data tersensor tipe II menggunakan data historis dari tabel CSO 1941 dimana jumlah total penelitian n berjumlah 1.000.000, waktu hidup t tahun, dan r sensor berupa r dimana pengamatan akan berhenti pada saat t_{r}. Diperoleh hasil bahwa, estimasi masksimal dengan ukuran sampel (n) terjadi pada estimator Weibull dengan $t_{\mathrm{r}} =$ 85 diperoleh $\widehat{\alpha} = -2{,}41\text{x}10^{-12}$ dan $\widehat{\beta} = 5$.

DAFTAR ISI

DAFTAR TABEL

Tabel Halaman

DAFTAR GAMBAR

DAFTAR LAMPIRAN

Lampiran Halaman

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Statistika inferensi merupakan salah satu cabang statistika yang berguna untuk menaksir parameter. Penaksiran dapat diartikan sebagai dugaan atau perkiraan atas sesuatu yang akan terjadi dalam kondisi tidak pasti, sedangkan parameter adalah nilai, cirri dari populasi. Dengan demikian penaksiran parameter merupakan suatu metode yang digunakan untuk memprediksi karakteristik suatu populasi berdasarkan sampel yang diambil.

Penaksiran parameter ada dua macam, yakni penaksiran titik dan penaksiran interval. Penaksiran titik diartikan sebagai penaksiran dari sebuah parameter populasi yang dinyatakan oleh sebuah bilangan tunggal. Sedangkan penaksir interval adalah penaksiran dari parameter populasi yang dinyatakan oleh dua bilangan diantara posisi parameternya. Penaksiran interval mengindikasikan tingkat kepresisian, atau akurasi dari sebuah panaksiran sehingga penaksiran interval akan dianggap semakin baik jika mendekati estimasi titik (Murray & Larry, 1999).

Perkiraan atau yang disebut *estimasi* adalah sebuah pemilihan yang unik untuk sebuah nilai parameter populasi yang tidak diketahui. Lebih jelasnya, jika X sebuah variabel random dengan distribusi probabilitas $f(x)$, mempunyai parameter θ yang tidak diketahui, dan jika $X_1, X_2, \dots, X_n$ sebuah sampel random yang besarnya n dari X, maka statistik $\bar{\theta} = h(X_1, X_2, \dots, X_n)$ yang berhubungan dengan θ disebut *estimator* θ (William W dan Dauglas C, 1990). Ada beberapa perbedaan utama estimator tunggal untuk suatu parameter. Misalnya, jika ingin memperkirakan rata-rata sebuah variabel random, maka harus memperhatikan salah satu rata-rata sampel, median sampel, atau mungkin observasi yang paling kecil dalam sampel sebagai estimator tunggal. Agar dapat menentukan estimator tunggal, sebuah parameter tertentu adalah satu-satunya yang paling terbaik untuk digunakan, perlu menguji sifat-sifatnya secara statistik dan mengembangkan beberapa kriteria untuk perbandingan estimator.

Berkaitan dengan dimensi fraktal, perkiraan yang tepat diterapkan yaitu metode *Maximum Likelihood Estimation* (MLE) yang merupakan suatu model estimator yang dapat mengatasi kelemahan metode Demografi konvensional, karena MLE θ adalah nilai θ yang memaksimumkan fungsi likelihood L(θ). Yang penting di dalam estimator *maximum likelihood* adalah nilai $\hat{\theta}$ yang memaksimumkan probabilitas kejadian hasil sampel. Disisi lain, MLE merupakan metode estimasi yang sangat populer dan digunakan secara luas untuk mengestimasi parameter suatu distribusi. Selain itu, Scot dan Nowak (2004) menunjukan bahwa terdapat banyak kelebihan yang dimiliki metode MLE, seperti sifat invariant, seringkali lebih sederhana dan mudah dalam perhitungan, dan mempunyai sifat konsistensi dan efisiensi asimtotik.

Untuk mengestimasi $\hat{\theta}$ pada metode MLE ada beberapa cara. Dalam hal ini, $\bar{\theta}$ ditentukan melalui pendekatan suatu distribusi, misalnya distribusi Weibull. Suatu variabel random X, berdistribusi Weibull dengan parameter bentuk β dan parameter skala α, jika fungsi densitasnya adalah:

$$f(t) = \begin{cases} 0 & t < 0 \\ \alpha\beta t^{\beta-1} e^{-\alpha t^{\beta}} & t \; yang \; lain \end{cases}$$

(1.1)

Distribusi Weibull digunakan secara luas bahwa notasi yang singkat $X \sim f(\alpha, \beta)$ sering digunakan untuk menunjukan bahwa variabel random X adalah berdistribusi Weibull dengan 2 parameter β dan α.

Untuk memperoleh $\hat{\theta}$ pada metode MLE melalui model distribusi yang diperlukan adalah nilai parameter θ (indeks stabilitas/*index of stability*) karena parameter θ menyatakan ketebalan ekor dari distribusi yaitu nilai α yang lebih kecil menerangkan ekor yang lebih tebal pada distribusi. Salah satu estimasi yang populer untuk indeks θ adalah dengan *Maximum Likelihood Estimation*. Parameter θ yang ditaksir menggunakan *Maximum Likelihood Estimation* selanjutnya disebut dengan fungsi $L(\theta)$.

Berdasarkan uraian tersebut penulis mengambil judul skripsi, yaitu "Estimasi Parameter Distribusi Weibull Tersensor-II Menggunakan MLE"

1.2 Rumusan Masalah

Dari latar belakang di atas rumusan masalah yang timbul dalam penulisan skripsi ini sebagai berikut.

1) Bagaimana metode estimasi data berdistribusi weibull tersensor tipe II menggunakan *Maximum Likelihood Estimation* (MLE)?
2) Bagaimana menentukan fungsi *Likelihood* $L(\alpha, \beta)$ dari data berdistribusi Weibull tersensor tipe II?

3) Bagaimana cara mendapatkan parameter α dan β menggunakan metode *Maximum Likelihood Estimation* (MLE)?

1.3 Pembatasan Masalah

Pembatasan masalah dalam penelitian ini adalah sebagai berikut:

1) Estimasi menggunakan metode *Maximum Likelihood Estimation* pada data tersensor tipe II berdistribusi Weibull.
2) Fungsi *likelihood* $L(\theta)$ ditentukan dari hubungan antara metode *Maximum Likelihood Estimation* pada data tersensor tipe II berdistribusi Weibull.
3) Estimasi mortalitas dengan metode *Maximum Likelihood Estimation* (MLE) dari data tersensor tipe II berdistribusi Weibull.

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1) Menentukan fungsi *Likelihood* $L(\alpha, \beta)$ pada data tersensor tipe II berdistribusi Weibull.
2) Menentukan nilai dari estimasi parameter α dan β dengan metode *Maximum Likelihood Estimation* (MLE) pada data tersensor tipe II berdistribusi Weibull.
3) Menjabarkan estimator α dan β pada tabel mortalitas menggunakan metode *Maximum Likelihood Estimation* (MLE) pada data tersensor tipe II berdistribusi Weibull.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini diharapkan dapat memberikan kontribusi dalam pengenalan dan pemahaman tentang Estimasi menggunakan metode *Maximum Likelihood Estimation* (MLE) pada data tersensor tipe II berdistribusi Weibull, karena model distribusi Weibull sendiri memberikan ruang yang lebih luas dalam penggunaannya sehingga mampu memberikan solusi lain dalam penyelesaian suatu masalah statistika dalam kehidupan nyata.

1.6 Sistematika Penulisan

Skripsi ini terbagi atas lima bab. Bab 1 berisi pendahuluan. Bab 2 landasan teori yang berisi teori konsep-konsep dasar probabilitas dan statistika serta karakteristik distribusi yang digunakan dalam pembahasan bab selanjutnya. Bab 3 berisi metodologi penelitian. Bab 4 berisi pembahasan menentukan estimator menggunakan metode *Maximum Likelihood Estimation* (MLE) pada data tersensor tipe II berdistribusi Weibull. Bab 5 berisi simpulan yang diperoleh dari pembahasan dalam bab 4 disertai saran.

BAB 2

LANDASAN TEORI

Dalam bab ini dipaparkan berbagai teori pendukung berkaitan dengan mortalitas, fungsi kepadatan peluang, fungsi distribusi, distribusi Weibull, fungsi survival, fungsi hazard, data tersensor dan metode estimasi maksimum likelihood.

2.1 Mortalitas

Mortalitas atau kematian merupakan salah satu di antara tiga komponen demografi yang dapat mempengaruhi perubahan penduduk. Meskipun manusia tidak dapat melangkahi kehendak Tuhan untuk memastikan kapan seseorang meninggal tetapi bukan rahasia cepat atau lambat setiap orang akan meninggal. Jika menggunakan asumsi bahwa kondisi yang menyebabkan terjadinya kematian di waktu yang akan datang tidak akan berbeda secara radikal dengan kondisi pada beberapa waktu yang lampau hingga saat sekarang, maka dapat dibuat prediksi secara umum berkenaan dengan kematian yang akan datang. Untuk dapat memprediksi sebagai contoh, beberapa orang terpilih secara acak dari sekelompok besar orang dapat diperkirakan akan meninggal dalam suatu periode waktu tertentu. Meskipun tidak dapat di tentukan dengan pasti mana individu yang akan meninggal.

Angka kematian atau *death rate* adalah suatu nilai probabilitas bahwa seseorang yang telah mencapai usia tertentu akan meninggal dalam waktu satu tahun. Tinggi rendahnya angka kematian dipengaruhi oleh berbagai faktor, misalnya struktur umur, jenis kelamin, jenis pekerjaan, status sosial ekonomi, keadaan lingkungan dan sebagainya (Dasar-dasar Demografi, 2004). Beberapa ukuran angka kematian yang sederhana antara lain:

2.1.1 Angka Kematian Kasar (*Annual Crude Death Rate)*

Angka kematian kasar per tahun didefinisakan sebagai jumlah kematian dalam satu tahun dibagi jumlah populasi pada pertengahan tahun tanggal 1 juli dari tahun yang bersangkutan, kemudian dikali dengan k. Angka ini menunjukan jumlah orang yang meninggal per k penduduk. Secara konvensional dinyatakan dengan rumus:

$$CDR = \frac{D}{p}.k \tag{2.1}$$

dengan D : Jumlah kematian pada tahun X

p : Jumlah penduduk pada pertengahan tahun X

k : 1.000

2.1.2 Angka Kematian Menurut Umur (*Age Specific Death Rate)*

Angka kematian menurut umur didefinisikan sebagai jumlah kematian kelompok umur tertentu pada pertengahan tahun tanggal 1 juli dibagi jumlah populasi, dikalikan dengan k. Sehingga dalam bentuk rumus dinyatakan dengan:

$$ASDR_i = \frac{D_i}{p_i}.k \tag{2.2}$$

dengan D_i : Jumlah kematian dari populasi kelompok umur i tahun X

p_i : Jumlah populasi kelompok umur i pada pertengahan tahun X

k : 1.000

2.1.3 Proporsi Rasio Kematian (*Propotional Mortality Ratio)*

Proporsi rasio kematian didefiniskan sebagai jumlah kematian yang disebabkan kasus khusus dalam setahun per jumlah kematian total pada tahun X, dikali dengan k. Sehingga dalam bentuk rumus dinyatakan dengan:

$$PMR = \frac{D_p}{D}.k \tag{2.3}$$

dengan D_p : Jumlah rasio kematian dari penduduk tahun X

D : Jumlah total kematian pada tahun X

k : 1000

2.1.4 Rasio Kematian Melahirkan (*Maternal Mortality Ratio)*

Proporsi rasio kematian melahirkan didefinisikan sebagai jumlah kematian yang diperuntukan untuk kasus kematian melahirkan tahun X per jumlah melahirkan hidup pada tahun X, dikali dengan k. sehingga dalam bentuk rumus dinyatakan dengan:

$$MMR = \frac{D_m}{p_m}.k \tag{2.4}$$

dengan D_m : Jumlah kematian dari ibu melahirkan kelompok pada tahun X

p_i : Jumlah ibu melahirkan hidup pada tahun X

k : 1.000

2.1.5 Angka Kematian Bayi (*Infant Mortality Ratio*)

Angka kematian bayi didefinisikan jumlah kematian bayi dibawah 1 tahun per k kelahiran dalam satu tahun. Angka kematian bayi dapat dirumuskan sebagai berikut:

$$IMR = \frac{D_o}{B}.k \tag{2.5}$$

dengan D_0 : Jumlah kematian bayi dibawah usia 1 tahun pada tahun X

B : Jumlah kelahiran hidup pada tahun X

k : 1.000

2.2 Tabel Kehidupan (*Life table*)

Tabel kehidupan yang sering disebut juga dengan tabel mortalitas adalah sebuah alat analisis mortalitas yang sering digunakan dan hasilnya cukup memuaskan para ahli demografi. Sebuah tabel mortalitas dikontruksi secara matematis untuk memberikan dekripsi secara lengkap mengenai angka kematian dan harapan hidup serta menunjukan pola mortalitas dari sekumpulan orang yang dilahirkan pada waktu yang bersamaan atau sering disebut kohor, berdasarkan usia yang telah dicapainya. Dalam perindustrian asuransi jiwa, para ahli aktuaria berfokus pada analisis mortalitas serta penyusunan tabel kehidupan untuk diaplikasikan pada perhitungan premi serta resiko guna memperoleh keuntungan bagi pengusaha (Daykin: 1995).

Tabel kehidupan merupakan sebuah matrik persegi panjang yang menunjukan perubahan masing-masing fungsi didalamnya melalui kolom, berdasarkan usia yang ditunjukan melalui baris. Pada dasarnya, terdapat dua jenis tabel kehidupan yang berbeda menurut panjang interval yang digunakan untuk menampilkan umur.

Pada umumnya komponen atau fungsi-fungsi standar yang terdapat dalam suatu tabel kehidupan adalah sebagai berikut:

a. Interval umur/*Age Interval* $[x \text{ to } (x+n)]$

   Interval umur adalah selisih antara nilai umur x dengan umur $x+n$ berikutnya. Pada *abridged life table* kelompok umur disusun dengan menggunakan interval 4 tahun untuk umur $1-4$ tahun, dan seterusnya sampai pada $\widehat{\omega}$ yaitu umur tertinggi yang membatasi umur pada tabel mortalitas. Kelompok umur pertama dan kedua tidak menggunakan interval 5 tahun dikarenakan khusus dan pentingnya angka kematian bayi (*Epidemiological Bulletin*).

b. Angka kematian antara umur tepat x dan $x+n$/Age Specific Death Rate and Age $x+n$ $({}_n m_x)$

   Simbol ${}_n m_x$ dinotasikan sebagai rata-rata kematian pada kelompok umur x tahun dan $x+n$ tahun, atau banyaknya kematian per jumlah "*person years*" hidup antara umur x tahun dan $x+n$ tahun, sehingga dirumuskan sebagai berikut:

$$ {}_n m_x = \frac{{}_n d_x}{{}_n l_x} \tag{2.6} $$

c. Syarat koreksi/*Correction Term* (${}_na_x$)

Diantara angka kematian bayi, kebanyakan kematian terjadi pada awal tahun pertama. Kolom ${}_na_x$ menunjukan rata-rata tahun hidup orang yang meninggal sebelum mencapai $x + n$.

d. Probabilitas kematian antara umur tepat x dan $x + n$ tahun/*Corrected Death rate* (${}_n\hat{q}_x$)

Nilai probabilitas ini dihitung berdasarkan angka kematian pada kelompok umur tertentu, sehingga dirumuskan sebagai berikut:

$$ {}_nq_x = \frac{n.\ {}_nM_x}{1 + n(1 - {}_na_x)\ {}_nM_x} \tag{2.7} $$

${}_na_x$ merupakan angka rata-rata tahun yang dijalani oleh orang yang berumur x yang meninggal pada umur antara x dan $x + n$ tahun, meski demikian tidak dicantumkan dalam tabel mortalitas tetapi nilai ini sangat penting dalam perhitungan. Nilai ${}_na_x$ diperoleh melalui fraksi dari interval terakhir kelompok umur, biasanya digunakan ${}_na_x = \frac{n}{2}$, sehingga nilai ${}_na_x$ untuk kematian pada kelompok umur dengan interval 5 tahun adalah 2,5 tahun. Namun untuk kelompok 0 dan $1 - 4$ tahun digunakan nilai ${}_na_x$ berdasarkan *Infant Mortality Rate* (IMR). Jika nilai l_x dan l_{x+n} terlebih dahulu diketahui, maka dapat digunakan rumus sebagai berikut:

$$ {}_nq_x = \frac{l_x - l_{x+n}}{l_x} = \frac{{}_nd_x}{l_x} \tag{2.8} $$

Jika kelompok umur terakhir pada tabel mortalitas menggunakan interval tak terhingga maka probabilitas kematian ${}_{\infty}q_{\omega} = 1$.

e. Jumlah orang yang berhasil mencapai umur tepat x tahun/*Number Living at beginning of Age Interval* (l_x)

l_x digunakan untuk memperkirakan proporsi dari individu yang bertahan hidup hingga umur awal dari kelompok umurnya. Seiring bertambahnya umur maka jumlah orang yang bertahan hidup pun berkurang. Secara umum nilai ekspektasi jumlah orang yang hidup pada umur x dirumuskan:

$$l_{x+n} = l_x(1 - {}_nq_x) \tag{2.9}$$

dimulai dari l_0 disebut sebagai radik tabel mortalitas, yaitu jumlah orang saat tepat lahir yang akan diikuti sampai semua orang tersebut meninggal. Nilainya ditentukan secara sebarang karena besar kecilnya tidak akan mempengaruhi interpretasi tabel mortalitas, biasanya ditentukan 100.000 atau 1.000.000. tentunya tabel mortalitas dapat dimulai dari umur berapapun (α) sehingga radik dituliskan dengan l_α.

f. Jumlah Kematian diantara Interval Umur/*Number Dying During Age Interval* (${}_nd_x$)

${}_nd_x$ menunjukan banyaknya kematian l_x yang terjadi antara umur x hingga umur $x + n$ tahun. Angka ini biasanya hanya digunakan dalam perhitungan dan tidak selalu ditampilkan pada tabel mortalitas.

$${}_nd_x = l_x - l_{x+n} \tag{2.10}$$

g. Jumlah tahun umur hidup dalam interval x/*Person Years Lived in Interval* $({}_nL_x)$

${}_nL_x$ merupakan ekspektasi jumlah dalam "tahun orang" (*person years*) dari l_x orang yang berumur tepat x yang menjalani hidup hingga umur $x+n$. Seorang anggota *kohort* yang menjalani hidup mulai dari awal umur x hingga umur $x+n$ berarti menjalani 2 "tahun orang hidup". Besarnya nilai ${}_nL_x$ diperkirakan dapat dilakukan dengan menggunakan rumus:

$$ {}_nL_x = \frac{n}{2}(l_x + l_{x+n}) \tag{2.11} $$

Nilai ${}_nL_x$ diperkirakan dengan menggunakan nilai ${}_na_x$ dan ${}_nd_x$, maka dapat dirumuskan:

$$ {}_nL_x = n.\, l_{x+n} + {}_na_x .\, {}_nd_x \tag{2.12} $$

h. Jumlah keseluruhan tahun orang hidup/*Total Number of Person Years* (T_x)

T_x adalah perkiraan jumlah "*person years*" yang dijalani hidup oleh l_x orang sejak umur tepat x sampai semuanya meninggal.

$$ T_x = \sum_{n=1}^{\infty} {}_nL_x \tag{2.13} $$

i. Angka harapan hidup/*Expectation of Life* $(\hat{e}_x)$

$\hat{e}_x$ menunjukan angka perkiraan rata-rata tahun hidup yang masih akan dijalani oleh anggota *kohort* setelah ia mencapai ulang tahunnya yang ke-x, dengan membagi "*person years*" yang dijalani oleh l_x orang sejak umur x, maka dapat diperoleh perkiraan angka harapan hidup. Secara umum dirumuskan sebagai berikut:

$$\hat{e}_x = \frac{1}{l_x}\sum_{n=1}^{\infty} {}_n l_x \tag{2.14}$$

2.3 Konsep Dasar Peluang

Pada dasarnya statistik berkaitan dengan penyajian dan penafsiran hasil yang berkemungkinan (hasil yang belum dapat ditentukan sebelumnya) yang muncul dalam penelitian yang dirancang sebelumnya atau yang muncul dalam penelitian ilmiah. Para statistisi berurusan dengan pencacahan atau pengukuran karakteristik suatu objek kajian yang hasilnya berbentuk bilangan. Pekerjaan seperti ini biasanya disebut percobaan acak (Abadyo dan Hendro permadi, 2005).

Himpunan semua hasil yang mungkin dari suatu percobaan acak disebut ruang sampel dan dinyatakan dengan lambang S. Suatu kejadian adalah himpunan bagian dari ruang sampel (Bain dan Engelhardt, 1991).

Ruang nol atau ruang kosong atau himpunan kosong adalah himpunan bagian ruang sampel yang tidak mengandung satu pun anggota. Kejadian seperti ini dinyatakan dengan lambang ∅ (Walpole, 1995).

Menurut Bain dan Engelhard (1992), untuk sebuah percobaan, S merupakan ruang sampel dan $A_1, A_2, A_3, \ldots$ merepresentasikan kejadian-kejadian yang mungkin. Hubungan fungsi yang menghubungkan nilai $P(A)$ dengan setiap kejadian A disebut himpunan fungsi peluang, dan $P(A)$ merupakan peluang dari A, jika memenuhi keadaan sebagai berikut:

i. $0 \leq P(A)$ untuk setiap A

ii. $P(S) = 1$

iii. $P(\bigcup_{i=1}^{\infty} A_1) = \sum_{i=1}^{\infty} P(A_i)$

dan jika $A_1, A_2, \ldots, A_i$ merupakan kejadian-kejadian yang saling asing

Definisi 2.1 *himpunan dari semua kemungkinan hasil (outcome) dari suatu percobaan disebut ruang contoh (sample space), dinotasikan dengan S.* Maka:

Ilustrasi 2.1 Jika sebuah dadu sisi enam dilempar, maka ruang contoh S adalah suatu himpunan yang memiliki 6 unsur yaitu $\{1,2, \ldots ,6\}$

Ilustrasi 2.2 Perhatikan pelemparan dua dadu sisi enam, ruang contoh yang mungkin adalah $S = \{(1,1), (1,2), (1,3), \ldots, (6,6)\}$. Dalam hal ini (i, j) berarti: i= mata dadu pertama dan j= mata dadu kedua yang muncul.

Ruang contoh bersifat tidak unik, tergantung dari sudut pandang, keperluan dan tujuan atau permasalahan.

Definisi 2.2 *kejadian (event) adalah himpunan bagian dari ruang contoh S.*

Medan- σ /Medan-Borel adalah suatu himpunan β yang anggotanya adalah kejadian-kejadian dalam ruang contoh S (kejadian) yang memenuhi tiga syarat berikut:

i. $\emptyset \in \beta$

ii. Jika $A \in \beta$ maka $A^C \in \beta$

iii. Jika $A_1, A_2, \ldots \in \beta$ maka $A_1 \cup A_2 \cup \ldots \in \beta$

Ilustrasi medan-σ:

i. $\beta = \{\emptyset, S\}$

ii. $\beta = \{\emptyset, A, A^C, S\}$ bila $A \subset S$

Definisi 2.3 *ukuran peluang P adalah suatu fungsi dari medan-σ ke selang tertutup* $[0,1](P: \beta \rightarrow [0,1])$ *yang memenuhi tiga syarat berikut:*

i. $P(A) \geq 0$, untuk setiap $A \in \beta$

ii. $P(S) = 1$

iii. Jika $A_1, A_2, .. \in \beta$ adalah himpunan yang saling lepas, yaitu $A_i \cap A_j = \emptyset$ untuk setiap pasangan i, j dengan $i \neq j$ maka:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \tag{2.15}$$

Teorema 2.1 *misalkan A dan B adalah kejadian dalam ruang contoh S dan A^C menyatakan komplemen dari A,* maka:

i. $P(A^C) = 1 - P(A)$

ii. Jika $A \subseteq B$ maka $P(A) \leq P(B)$

iii. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Bukti.

Perhatikan bahwa $A \cup A^C = S$ dan $A \cap A^C = \emptyset$.

Karena A dan A^C saling lepas, maka

$$P(A \cup A^C) = P(A) + P(A^C) = 1$$

Jadi

$$P(A^C) = 1 - P(A) \tag{2.16}$$

2.3.1 Peluang Bersyarat

Definisi 2.4 *peluang kejadian A dengan syarat bahwa kejadian B telah diketahui terjadi adalah*

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad P(B) > 0$$

(2.17)

Teorema 2.2

a. *Untuk sebarang kejadian A dan B berlaku*

$$P(A) = P(B)P(A|B) + P(B^C)P(A|B^C) \quad P(B) > 0$$

b. *Secara umum misalkan ada $B_1, B_2, .., B_n$ adalah partisi yang bersifat saling lepas dari S,* maka

$$P(A) = P(B_1)P(A|B_1) + .. + P(B_n)P(A|B_n)$$
$$= \sum_{i=1}^{n} P(B_i)P(A|B_i)$$

(2.18)

2.3.2 Kejadian Bebas

Jika A adalah suatu kejadian, maka adanya keterangan tentang suatu kejadian lain, misal kejadian B, dapat memperkecil atau memperbesar atau tidak mengubah besarnya peluang kejadian A. Jika besarnya peluang kejadian A tidak berubah karena adanya keterangan bahwa kejadian B telah terjadi, maka A dan B adalah dua kejadian yang saling bebas.

Definisi 2.5 *kejadian A dan B disebut dua kejadian yang saling bebas jika dan hanya jika* $P(A \cap B) = P(A).P(B)$

Teorema 2.3 *jika A dan B adalah dua kejadian bebas*, maka:

i. A dan B^C adalah dua kejadian bebas

ii. A^C dan B juga dua kejadian bebas

iii. A^C dan B^C juga dua kejadian bebas

Bukti.

akan ditunjukan bahwa $P(A \cap B^C) = P(A).P(B^C)$

karena $A = (A \cap B) \cup (A \cap B^C)$ dan $(A \cap B) \cap (A \cap B^C) = \emptyset$

maka $P(A) = P(A \cap B) + P(A \cap B^C)$

jadi $P(A \cap B^C) = P(A) - P(A \cap B^C)$

$$= P(A) - P(A)P(B)$$

$$= P(A)\big(1 - P(B)\big)$$

$$= P(A)P(B^C)$$

(2.19)

Jika kejadian A dan B bebas, maka kejadian bersyaratnya tidak merubah nilai peluang.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A) \quad P(B) > 0$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A)P(B)}{P(A)} = P(B) \quad P(A) > 0$$

(2.20)

2.4 Peubah Acak

Definisi 2.6 *peubah acak adalah suatu fungsi yang menghubungkan sebuah bilangan real dengan setiap unsur di dalam ruang contoh* (Walpole & Dkk, 2003)

Peubah acak dilambangkan dengan huruf capital X dan huruf kecilnya dalam hal ini x, untuk menyatakan salah satu diantara nilai-nilainya. Dengan demikian suatu bilangan X merupakan ukuran dari karakteristik yang diletakkan pada setiap kejadian dasar dari ruang contohnya. Peubah acak diklasifikasikan menjadi 2 macam, yaitu peubah acak diskret dan peubah acak kontinu (Wibisono, 2005).

Definisi 2.7 *X disebut peubah acak diskret bila X peubah acak yang hanya mendapat nilai berhingga atau banyaknya terbilang* (Dudewicz & Mishra, 1995).

2.8.1 Distribusi Peubah Acak Diskrit

Seringkali untuk memudahkan suatu perhitungan semua probabilitas peubah acak dinyatakan dalam fungsi nilai-nilai X seperti $f(x) = P(X = x)$. Pada peubah acak diskrit, setiap nilainya dikaitkan dengan probabilitas. Himpunan pasangan berurutan $[x, f(x)]$ disebut distribusi probabilitas peubah acak X. Sebuah distribusi yang mencantumkan semua kemungkinan nilai peubah acak diskrit berikut probabilitasnya disebut probabilitas diskrit (Wibisono, 2005: 224).

Suatu peubah acak diskrit dapat dinyatakan sebagai:

$$F(x) = \sum px_x(x) \tag{2.21}$$

Definisi 2.8 *himpunan pasangan terurut $\left(x, f(x)\right)$ merupakan suatu fungsi peluang, fungsi massa peluang, atau distribusi peluang acak diskrit X bils, untuk setiap kemungkinan hasil x:*

i. $f(x) \geq 0$

ii. $\sum f(x) = 1$

iii. $P(X = x) = f(x)$

(Walpole & Myers, 1995)

Definisi 2.9 *jika peubah X dapat menerima suatu himpunan diskrit dari nilai-nilai $X_1, X_2, \ldots, X_n$ dengan probabilitas masing-masing $P_1, P_2, \ldots, P_n$, dimana $P_1 + P_2 + \ldots + P_n = 1$, maka fungsi $f(X)$ yang mempunyai nilai masing-masing $P_1, P_2, \ldots, P_i$ untuk $X_1, X_2, \ldots, X_i$ disebut fungsi probabilitas. Sehingga dapat dituliskan dengam $f(X) = P(X = X_i)$, yaitu probabilitas P nilai peubah X ke-i (yaitu X_i) sama dengan $f(X)$* (Turmudi & Harini, 2008).

2.8.2 Distribusi Peubah Acak Kontinu

Distribusi probilitas bagi peubah acak kontinu tidak dapat disajikan dalam bentuk tabel, akan tetapi distribusinya dapat dinyatakan dalam persamaan yang merupakan fungsi nilai-nilai peubah acak kontinu dan digambarkan dalam bentuk kurva (Wibisono, 2005).

Suatu peubah acak kontinu dapat dinyatakan sebagai:

$$f(x) = \int_{-\infty}^{\infty} f_x(x)\, dx \tag{2.22}$$

Definisi 2.10 *fungsi $f(x)$ adalah fungsi padat peluang peubah acak kontinu X, yang didefinisikan atas himpunan semua bilangan real R*, bila:

i. $f(x) \geq 0 \in$ *untuk semua* $x \in R$

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

iii. $P(a < x < b) = \int_a^b f(x)\, dx$

(Walpole & Myers, 1995)

2.5 Ekspektasi Matematika

Konsep ekspektasi matematika (nilai harapan secara matematik) dalam statistik sangat besar manfaatnya. Selain digunakan untuk pengembangan dalam statistik lanjutan dan terapan dibidang lain, juga sebagai konsep dasar untuk mendefinisikan atau membangun ukuran-ukuran dalam statistik, seperti rata-rata, varian, koefisien, korelasi.

2.5.1 Pengertian Ekspektasi Matematika

Definisi 2.11 *misalkan p,a X dengan f.k.p $f(x)$ fungsi atau bentuk dalam X. Ekspektasi matematik atau nilai harapan dari $U(x)$, ditulis dengan $E[u(x)]$ dan*

i. *Untuk X diskrit*

$$E[u(x)] = \sum_{x\in} U(x)f(x) = \sum_{x\in\Re} u(x).P[X = x]$$

(2.23)

ii. *Untuk X kontinu*

$$E[u(x)] = \int_{-\infty}^{\infty} U(x)f(x)\,dx$$

(2.24)

Definisi 2.12 *misalkan $f(x,y)$ fkp bersama dari peubah acak X dan Y, dan $U(X,Y)$ fungsi dalam X dan Y. Ekspektasi matematika dari $U(X,Y)$ ditulis dengan $E[U,XY]$, dan*

i. *Untuk X,Y diskrit*

$$E[U(X,Y)] = \sum_{x\in\Re}\sum_{y\in\Re} u(x,y)f(x,y)$$

$$= \sum_{x \in \Re} \sum_{y \in \Re} u(x, y) P(X = x, Y = y)$$

(2.25)

ii. *Untuk X, Y kontinu*

$$E[U(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x, y) f(x, y)\, dx\, dy$$

(2.26)

Definisi 2.13 *misalkan $f(x_1, x_2, .., x_n)$ fkp bersama dari $x_1, x_2, .., x_n$ dan $u(x_1, x_2, .., x_n)$ fungsi dalam $x_1, x_2, ...$ dan x_n. Ekspektasi dari $u(x_1, x_2, .., x_n)$, dituliskan dengan $E[u(x_1, x_2, .., x_n)]$, dan*

i. *Untuk X, Y diskrit*

$$E[u(x_1, x_2, .., x_n)] = \sum_{x_1} \sum_{x_2} \dots \sum_{x_n} u(x_1, x_2, .., x_n) f(x_1, x_2, .., x_n)$$

(2.27)

ii. *Untuk X, Y kontinu*

$$E[u(x_1, x_2, .., x_n)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x_1, x_2, .., x_n), f(x_1, x_2, .., x_n)$$

(2.28)

Sifat-sifat Ekspektasi:

i. $E[k] = k, \forall k$ konstanta real

ii. $E[k, u(X)] = kE[u(X)], \forall k$ konstanta real

iii. $E[\sum_{l=1}^{N} k_i u_i X] = \sum_{i=1}^{n} k_i E[u_i X], \forall k_1, k_2, \dots, k_n$ konstanta real

2.5.2 Rerata dan Varian

Jika X peubah acak, dengan $u(X) = X$, maka $E[u(X)] = E[X]$

Definisi 2.14 *jika $f(x)$ fkp dari peubah acak X*, maka:

i. Untuk X peubah acak diskrit, rerata dari X adalah

$$\mu = E[u(X)] = \sum_{x} xf(x) \tag{2.29}$$

ii. Untuk X peubah acak kontinu, rerata dari X adalah

$$\mu E[X] = \int_{-0}^{0} x, f(x)\, dx \tag{2.30}$$

Definisi 2.15 *jika $f(x)$ fkp dari peubah acak X*, maka:

i. Untuk X peubah acak diskrit

$$\sigma^2 = var(x) = E[(x-\mu)^2] = \cdots (x-\mu)^2 f(x) \tag{2.31}$$

ii. Untuk X peubah acak kontinu

$$\sigma^2 = var(x) = E[(x-\mu)^2] = \cdots (x-\mu)^2 f(x)\, dx$$

(2.32)

Teorema 2.4 $Var(X^2) = E[X^2] - (E[X])^2 = E[X^2] - \mu^2$

Bukti.

$Var(X) = E(x^2 - \mu)^2 - E(x^2) - 2\mu x + \mu^2$

$= E(x^2) - 2\mu E(x) + E(\mu^2)$

$= E(x^2) - \mu^2$

(2.33)

Sifat-sifat Rerata dan Varians

Jika μ dan σ^2 berturut-turut adalah rerata dan varians dari X, maka:

i. $E[(X-\mu)] = 0$

ii. $E[aX+b] = a\mu + b$, $\forall a, b$ konstanta real

iii. $Var(X+c) = Var[x] =$, $\sigma^2 \forall c$ konstanta real

iv. $Var(aX+b) = a^2 Var[X] = a^2\sigma^2$

Bukti.

Akan dibuktikan hanya bagian (*iv*) saja.

$Var(aX+b) = E[(aX+b)^2] - E^2[aX+b]$

$= E[aX^2 + 2abX + b^2] - E^2[a\mu + b]^2$

$$= a^2E\big[[X^2] - \mu^2\big] + 2ab(E[X] - \mu)$$

$$= a^2Var[X] + 2ab(0)$$

$$= a^2\sigma^2$$

(2.34)

2.5.3 Momen dan Fungsi Pembangkit Momen

Definisi 2.16

i. *Momen ke k dari peubah acak X dinotasikan dengan μ adalah ekspektasi dari X^k, $k = 1,2,3,\ldots$ ditulis*

$$\mu_k = E[X^k]$$

(2.35)

ii. *Momen sentral ke k sekitar rerata μ dari p,a X dinotasikan dengan*

$$\mu_k = E[(X - \mu)^k], k = 1,2,3,\ldots$$

(2.36)

Jadi jika $f(x)$ fkp dari p,a X maka:

i. Untuk X diskrit:

$$\mu_k = E[X^k] = \sum_x (x)^k f(x)$$

$$\mu_k = E[(X-\mu)^k] = \sum_x (X-\mu)^k f(x)$$

(2.37)

ii. Untuk X kontinu

$$\mu_k = E[X^k] = \int_{-\infty}^{\infty} xk.\, f(x)\, dx$$

(2.38)

2.6 Distribusi Diskrit

2.6.1 Distribusi Binomial

Suatu percobaan yang terdiri dari dua hasil yang mungkin, katakanlah A dan bukan $A(\bar{A})$ dengan $P(A) = P$ dan $P(\bar{A}) = 1 - P$, berharga tetap dalam setiap percobaan, maka percobaan itu dinamakan Bernoeli.

Jika percobaan Bernoeli dilakukan sebanyak N kali secara independen, x diantaranya menghasilkan kejadian A dan sisanya $(n - x)$ kejadian $\bar{A}$ (bukan A), maka peluang terjadinya A sebanyak $X = x$ kali dari n percobaan, dinyatakan sebagai berikut:

$$b(x,n,p): P(X = x) = \binom{n}{x} P^x (1-P)^{n-x}$$

(2.39)

dengan

$$\binom{n}{x} = \frac{n!}{x!\,(n-x)!}$$

(2.40)

Hubungan yang dinyatakan oleh rumus (2.6.1) di atas dinamakan distribusi Binomial. Disekitar Binomial $b(x, n, p)$, mempunyai parameter rata-rata $\mu = np$ dan varians $\sigma = npq$.

2.6.2 Distribusi Multinominal

Jika percobaan Binomial berkembang dengan memberikan lebih dari dua hasil yang mungkin, bukan hanya kategori sukses atau gagal, maka percobaan itu dinamakan multinomial.

Misalkan sebuah percobaan menghasilkan K buah hasil yang mungkin $E_1, E_2, \dots, E_k$ dengan peluang $P_1, P_2, \dots, P_k$, jika percobaan dilakukan sebanyak n kali, maka peluang akan terdapat x_1, kejadian E_1, x_2 kejadian E_2,…, x_k kejadian E_k diantara n ditentukan oleh distribusi multinomial sebagai berikut:

$$P(x_1, x_2, \dots, x_k) = \frac{N!}{x_1!\, x_2! \dots x_k!} P_1^2 . P_2^2 \dots P_k^k$$

(2.41)

2.6.3 Distribusi Poisson

Teorema 2.5 *Jika $X_1, X_2, \dots, X_k$ adalah variabel random yang didistribusikan secara bebas, masing-masing mempunyai distribusi Poisson dengan parameter $\alpha_i, i = 1,2,\dots,k$, dan $Y = X_1 + X_2 + \cdots + X_k$, maka Y mempunyai distribusi Poisson dengan parameter*

$$\alpha = \alpha_1 + \alpha_2 + \cdots + \alpha_k \tag{2.42}$$

Bukti.

Fungsi pembangkit momen X_i adalah

$$M_{x_i}(t) = e^{\alpha_i(e^t - 1)} \tag{2.43}$$

dan karena $M_Y(t) = M_{x_1}(t).M_{x_2}(t) \cdot \ldots \cdot M_{x_k}(t)$, maka

$$M_Y(t) = e^{(\alpha_1 + \alpha_2 + \cdots + \alpha_k)(e^t - 1)} \tag{2.44}$$

yang mana dikenal sebagai fungsi pembangkit momen dari sebuah variabel random poisson dengan parameter $\alpha = \alpha_1 + \alpha_2 + \cdots + \alpha_k$.

Sifat yang dihasilkan distribusi poisson ini manfaatnya banyak sekali. Secara sederhana, ini menyatakan bahwa jumlah variabel random Poisson yang bebas didistribusikan menurut distribusi Poisson tersebut.

2.7 Distribusi Kontinu

2.7.1 Distribusi Normal

Distribusi normal dalam beberapa hal merupakan dasar dari statistik. Sebuah variabel random X, disebut mempunyai sebuah distribusi normal dengan rata-rata $\mu(-\infty < \mu < \infty)$ dan varian $\sigma^2 > 0$, jika bentuk fungsi kepadatan peluangnya adalah

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(1/2)[(x-\mu)/\sigma]^2} \quad -\infty < x < \infty$$

(2.43)

Distribusi normal digunakan secara luas bahwa notasi yang singkat $X \sim N(\mu, \sigma^2)$ sering digunakan untuk menunjukan bahwa variabel random X adalah berdistribusi normal dengan rata-rata μ dan varian σ^2.

Beberapa sifat penting yang dimiliki distribusi normal:

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

ii. $f(x) \geq 0\ untuk\ seluruh\ x$

iii. $\lim_{x \to \infty} f(x) = 0$ dan $\lim_{x \to -\infty} f(x) = 0$

iv. $f[(x + \mu)] = f[-(x - \mu)]$ densitas tersebut adalah simetris di sekitar μ.

v. Nilai maksimum dari f terjadi pada $x = \mu$.

vi. Titik perubahan dari f terjadi pada $x = \mu \pm \sigma$.

Sifat (1) dapat ditunjukan sebagai berikut. Misalkan $y = (x - \mu)/\sigma$ dalam persamaan $(-)$ dan dinotasikan integral sebagai I. Maka,

$$I = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(1/2)y^2}\, dy$$

(2.44)

Akan dibuktikan bahwa

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

(2.45)

akan berisikan gambaran bahwa $I^2 = 1$, dan kemudian diambil kesimpulan bahwa $I = 1$, karena f harus bernilai positif. Misalkan variabel kedua yang berdistribusi normal Z, di punyai:

$$I^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(1/2)/y^2}\, dy - \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-(1/2)z^2}\, dz$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(1/2)(y^2+z^2)}\, dy\, dz$$

(2.46)

Dengan mengubah pada kordinat polar dengan transformasi dari variabel $y = r \sin\theta$ dan $z = r\cos\theta$, integral tersebut menjadi

$$I^2 = \frac{1}{2\pi} \int_{0}^{\infty} \int_{0}^{2\pi} re^{-(1/2)r^2}\, d\theta\, dr$$

$$= \int_{0}^{\infty} re^{-(1/2)r^2} dr = 1$$

(2.47)

Melengkapi pembuktian tersebut rata-rata dari distribusi normal dapat ditentukan dengan mudah karena

$$E(X) = \int_{-\infty}^{\infty} \frac{x}{\sigma\sqrt{2\pi}} e^{-(1/2)[(x-\mu)/\sigma]^2}$$

(2.48)

Misalkan $z = \frac{(x-\mu)}{\sigma}$, peroleh

$$E(X) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} (\mu + \sigma z) e^{-\frac{z^2}{2}} dz$$

$$= \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz + \sigma \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} z e^{-z^2/2} dz$$

(2.49)

Karena bagian integral pertama adalah sebuah densitas normal dengan $\mu = 0$ dan $\sigma^2 = 1$, nilai integral pertama adalah satu. Bagian integral kedua mempunyai nilai nol, yaitu,

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = -\frac{1}{\sqrt{2\pi}} e^{-z^2/2} \Big|_{-\infty}^{\infty} = 0$$

(2.50)

Sehingga

$$E(X) = \mu[1] + \sigma[0] = \mu$$

Untuk mendapatkan varian, maka harus dihitung

$$V(X) = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-(1/2)[x-\mu/\sigma]^2} dx$$

(2.51)

Dan misalkan $z = \frac{(x-\mu)}{\sigma}$ diperoleh

$$V(X) = \int_{-\infty}^{\infty} \sigma^2 z^2 \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

$$= \sigma^2 \left[\int_{-\infty}^{\infty} \frac{z^2}{\sqrt{2\pi}} e^{-z^2/2} dz \right]$$

$$= \sigma^2 \left[\frac{-ze^{-z^2/2}}{\sqrt{2\pi}} \bigg|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \right]$$

$$= \sigma^2 [0 + 1]$$

$$= \sigma^2$$

Sehingga

$$V(X) = \sigma^2 \tag{2.52}$$

Fungsi pembangkit momen untuk distribusi normal dapat ditunjukan menjadi

$$M_x(t) = e^{\left(t\mu + \sigma^2 t^2/2\right)} \tag{2.53}$$

2.7.2 Distribusi Log-logistik

Dalam statistika, distribusi log-logistik merupakan salah satu distribusi peluang kontinu untuk variabel random non-negatif. Distribusi ini digunakan dalam analisis tahan hidup sebagai model parametrik, misalnya untuk meneliti waktu penyembuhan penyakit. Distribusi log-logistik juga telah dikembangkan di bidang industri untuk menganalisis tahan hidup komponen dari suatu produk.

Variabel random T dikatakan mengikuti distribusi log-logistik dengan parameter γ dan parameter *shape* β, jika mempunyai fungsi kepadatan peluang:

$$f(t) = \frac{\left(\frac{\beta}{\gamma}\right)\left(\frac{t}{\gamma}\right)^{\beta-1}}{\left[1+\left(\frac{t}{\gamma}\right)^{\beta}\right]^2} \quad t, \gamma, \beta > 0$$

(2.54)

Untuk selanjutnya dinotasikan sebagai $T \sim L_L(\gamma, \beta)$. Nilai parameter *shape* yaitu β menyatakan suatu bentuk yang macam-macam dari kurva fungsi densitas yaitu naik, turun, atau mendatar, sehingga kondisi ini cocok digunakan untuk berbagai model data *survival*.

2.7.3 Distribusi Weibull

Distribusi Weibull merupakan salah satu dari distribusi yang dapat digunakan untuk memodelkan suatu fenomena kerusakan dengan *hazard rate* $h(t)$ tergantung pada usia pakai suatu alat/komponen tersebut. Distribusi Weibull yang digunakan dalam penelitian ini adalah distribusi Weibull dengan dua parameter yaitu α dan β.

Distribusi Weibull dua parameter dapat mereprensentasikan *hazard rate* yang meningkat dan konstan. Menurut jurnal Asadi (2011) tentang "*Estimation of the Weibull Distribution Based on Type-II Censored Sample"* fungsi kepadatan peluang distribusi Weibull dinyatakan sebagai berikut:

$$f(t) = \begin{cases} 0 & t < 0 \\ \alpha\beta t^{\beta-1} e^{-\alpha t^{\beta}} & t \text{ yang lain} \end{cases}$$

(2.55)

dengan t : waktu, $t > 0$

α : parameter skala, $\alpha > 0$

β : parameter bentuk, $\beta > 0$

2.8 Konsep Dasar Distribusi Survival

2.8.1 Fungsi Distribusi kumulatif

jika T merupakan variabel random dari waktu hidup suatu individu dalam interval $[0, \infty)$, maka fungsi distribusi kumulatif $F(t)$ untuk distribusi kontinu dengan fungsi kepadatan peluang $f(t)$ dinyatakan sebagai berikut (Lawless, 1982):

$$F(t) = P(T \leq t)$$

atau

$$F(t) = \int_{0}^{t} (x)\, dx \quad t > 0$$

(2.56)

2.8.2 Fungsi *Survivor*

Menurut Lawless (1982) fungsi *survivor* didefinisikan sebagai peluang suatu individu dapat bertahan hidup sampai waktu t. Jika T merupakan variabel random dari waktu hidup suatu individu dalam interval $[0, \infty)$, maka fungsi *survivor* $S(t)$ dapat dinyatakan dalam persamaan:

$$S(t) = P(T \geq t)$$

$$= \int_{t}^{\infty} f(x)\, dx$$

$$= [f(x)]_t^{\infty}$$

$$= f(\infty) - f(t)$$

$$= 1 - f(t)$$

(2.57)

Dengan demikian diperoleh persamaan yang menyatakan hubungan antara fungsi *survivor* dan fungsi distribusi kumulatif, yaitu:

$$S(t) = 1 - F(t)$$

(2.58)

2.8.3 Fungsi Hazard

Fungsi *hazard* menyatakan peluang kegagalan suatu komponen pada waktu t, jika diketahui bahwa komponen tersebut tetap hidup hingga waktu t. Menurut Lawless (1982) fungsi *hazard* adalah peluang suatu individu mati dalam interval waktu t sampai $t + \Delta t$, jika diketahui individu tersebut masih dapat bertahan hidup sampai dengan waktu t, yang dinyatakan sebagai berikut:

$$h(t) = \lim_{\Delta t \to 0} \left[\frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \right]$$

(2.59)

jika $f(t)$ adalah fungsi densitas peluang pada waktu t, maka diperoleh

$$h(t) = \lim_{\Delta t \to 0} \left[\frac{P\big((\mathrm{t} \leq \mathrm{T}) < (t + \Delta t | T \geq t)\big)}{\Delta t} \right]$$

$$= \lim_{\Delta t \to 0} \left[\frac{P\left[\left((\mathrm{t} \leq \mathrm{T}) < (\mathrm{t} + \Delta\, t)\right) \cap (\mathrm{T} \geq \mathrm{t})\right]}{P(T \geq t)\, \Delta\, t} \right]$$

$$= \lim_{\Delta t \to 0} \left[\frac{P\left(t \leq T < (t + \Delta\, t)\right)}{P(T \geq t)\, \Delta\, t} \right]$$

dari persamaan (2.56), (2.56), dan (2.56) maka

$$= \lim_{\Delta t \to 0} \left[\frac{1}{\Delta\, t} \cdot \frac{F(t + \Delta\, t) - F(t)}{1 - F(t)} \right]$$

$$= \lim_{\Delta t \to 0} \left[\frac{1}{S(t)} \cdot \frac{F(t + \Delta\, t) - F(t)}{\Delta\, t} \right]$$

$$= \frac{F'(t)}{S(t)}$$

$$h(t) = \frac{f(t)}{S(t)} \tag{2.60}$$

2.9 Data Tersensor

Dalam penelitian uji hidup, data waktu hidup dapat berbentuk data lengkap, data tersensor tipe I dan data tersensor tipe II. Pada pengambilan data menggunakan data lengkap, percobaan akan dihentikan jika semua komponen atau individu yang diteliti gagal atau mati (Lawless, 1982). Metode menggunakan data lengkap memerlukan waktu yang lama sehingga jarang digunakan.

Data tersensor adalah data yang diperoleh sebelum semua data teramati waktu hidupnya, sedangkan waktu pengamatan telah berakhir atau oleh sebab lain dihentikan. Data tersensor tipe I merupakan data uji hidup yang dihasilkan setelah penelitian berjalan selama waktu yang telah ditentukan. Sedangkan data tersensor tipe II merupakan data hasil penelitian dimana penelitian dihentikan setelah kematian atau kegagalan tertentu telah terjadi (Lawless, 1982).

Data tersensor tipe II merupakan data kematian atau kegagalan yang tidak lengkap (*incomplete mortality data*) yaitu data waktu kematian atau kegagalan dari r observasi terkecil dalam sampel random yang berukuran n dengan $1 < r < n$. Dalam suatu penelitian, penyensoran tipe II lebih sering digunakan, yaitu dalam uji hidup yang terdapat observasi sebanyak n, tetapi penelitian dihentikan ketika observasi mengalami kegagalan ke-r, sehingga dapat menghemat waktu dan biaya. Dalam penyensoran ini, r ditentukan terlebih dahulu sebelum data dikumpulkan.

2.10 Metode *Maximum Likelihood Estimation*

Metode maksimum likelihood adalah salah satu metode yang paling sering digunakan untuk mencari nilai estimasi dari suatu parameter. Fungsi kepadatan bersama (*joint density function*) dari n variabel random $X_1, X_2, \dots, X_n$ pada $x_1, x_2, \dots, x_n$ adalah $f(x_1, x_2, \dots, x_n \cdot \theta)$ disebut sebagai fungsi likelihood. Untuk $x_1, x_2, \dots, x_n$ yang tetap, fungsi likelihood adalah fungsi dari θ dan sering dinotasikan $L(\theta)$.

Definisi 2.17 *jika $X_1, X_2, \dots, X_n$ menyatakan sampel random dengan densitas peluang $f(x;\theta)$* maka:

$$L(\theta) = f(x_1;\theta) \dots f(x_n;\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

(2.61)

Misalkan $L(\theta) = f(x_1;\theta) \dots f(x_n;\theta)$, $\theta\epsilon\Omega$ adalah fungsi kepadatan bersama dari $X_1, X_2, \dots, X_n$. Untuk sekumpulan observasi yang diberikan $(x_1, x_2, \dots, x_n)$ suatu nilai $\hat{\theta}$ dan Ω sedemikian hingga $L(\theta)$ maksimum, disebut *Maximum Likelihood Estimation* (MLE) dari θ. Nilai $\hat{\theta}$ adalah nilai θ yang memenuhi:

$$f\left(x_1, x_2, \dots, x_n; \hat{\theta}\right) = \max_{\theta\epsilon\Omega} f(x_1, x_2, \dots, x_n; \theta) \tag{2.62}$$

Apabila Ω adalah interval terbuka dan jika $L(\theta)$ adalah differensiabel dan diasumsikan maksimum pada Ω maka MLE adalah solusi dari persamaan:

$$\frac{d}{d\theta} L(\theta) = 0 \tag{2.63}$$

Hal yang perlu diperhatikan , jika ternyata terdapat lebih dari satu solusi untuk persamaan (2.10.3), maka harus dilakukan perhitungan terhadap masing-masing solusi untuk memperoleh solusi yang memaksimumkan $L(\theta)$. Hal ini dilakukan dengan mencari nilai turunan kedua dari $L(\theta)$.

Definisi tentang fungsi likelihood dan estimasi kemungkinan maksimum dapat diterapkan dalam parameter-parameter tak diketahui yang lebih dari satu. Bila θ adalah parameter, katakanlah $\theta = (\theta_1, \theta_2, \dots, \theta_k)$, maka estimasi kemungkinan maksimumnya akan berupa persamaan simultan dengam penurunan parsial tiap-tiap parameternya.

$$\frac{\delta}{\delta\theta_j} \ln L(\theta_1, \theta_2, \dots, \theta_k) = 0 \tag{2.64}$$

untuk $j = 1,2, \dots, k$.

Persamaan di atas disebut persamaan kemungkinan maksimum (*Maximum Likelihood Equations*). Nilai $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k$ merupakan estimator bila persamaan kemungkinan maksimumnya memberikan nilai maksimum terhadap

$$L(\theta) = (\theta_1, \theta_2, \dots, \theta_k).$$

2.11 Uji Asumsi Klasik

Model regresi linear berganda (*multiple regression)* dapat disebut sebagai model yang baik jika model tersebut memenuhi Kriteria Blue (*Best Linear Unbiased Estimator*). Blue dapat dicapai bila memenuhi Asumsi Klasik. Sedikitnya terdapat 4 uji asumsi yang harus dilakukan terhadap suatu model regresi tersebut, yaitu:

2.11.1 Uji Normalitas

Cara yang sering digunakan dalam menentukan apakah suatu model berdistribusi normal atau tidak hanya dengan melihat histogram residual apakah memiliki bentuk seperti "lonceng" atau tidak. Cara ini menjadi fatal karena pengambilan keputusan data berdistribusi normal atau tidak hanya berpatok pada pengamatan gambar. Cara lain untuk menentukan data berdistribusi normal atau tidak dengan menggunakan rasio skewness dan rasio kurtosis.

Rasio skewness dan rasio kurtosis dapat dijadikan petunjuk apakah suatu data berdistribusi normal atau tidak. Rasio skewness adalah nilai skewness dibagi dengan standard error kurtosis. Sebagai pedoman, bila rasio kurtosis dan skewness berada diantara -2 hingga $+2$, maka distribusi data adalah normal (Santoso, 2000: 53).

2.11.2 Uji Autokorelasi

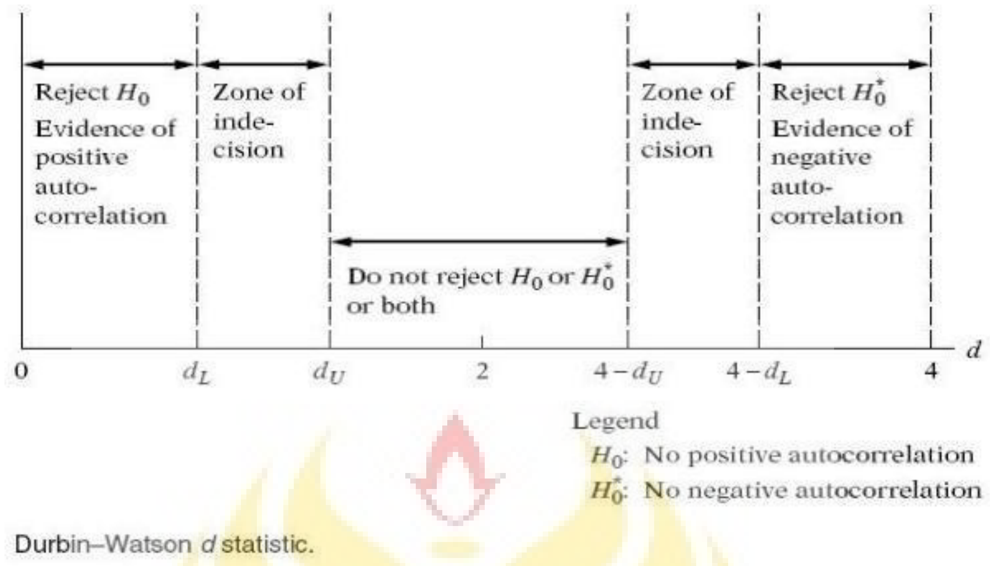
Ada beberapa cara dapat digunakan untuk mendeteksi ada tidaknya autokorelasi. Pertama, uji Durbin-Watson (DW Test). Uji ini hanya digunakan untuk autokorelasi tingkat satu (*first order autocoreelation*) dan mensyaratkan adanya intercept dalam model regresi dan tidak ada variabel lag di antara variabel penjelas. Hipotesis yang diuji adalah:

Ho: p = 0 (dibaca: hipotesis nolnya adalah tidak ada autokorelasi)

Ha: p ≠ 0 (dibaca: hipotesis alternatifnya adalah ada autokorelasi)

keputusan ada tidaknya autokorelasi adalah:

- Bila nilai DW berada di antara d_u sampai 4 - d_u maka koefisien autokorelasi sama dengan nol. Artinya, tidak ada autokorelasi.
- Bila nilai DW lebih kecil dari pada d_l, koefisien autokorelasi lebih besar dari pada nol. Artinya ada autokorelasi positif.
- Bila nilai DW terletak di antara d_l dan d_u, maka tidak dapat disimpulkan.
- Bila nilai DW lebih besar dari pada 4 - d_l, koefisien autokorelasi lebih besar dari pada nol. Artinya ada autokorelasi negatif.
- Bila nilai DW terletak di antara 4 - d_u dan 4 - d_l, maka tidak dapat disimpulkam.



Gambar 2.1 Pengambilan Keputusan Dengan DW Test

2.11.3 Uji Multikolinieritas

Ada banyak cara untuk menentukan apakah suatu model memiliki gejala Multikolinieritas, di antaranya, sebagai berikut:

- Uji VIF

  Cara ini sangat mudah, hanya melihat apakah nilai VIF untuk masing-masing variabel lebih besar dari 10 atau tidak. Bila nilai VIF lebih besar dari 10 maka di indikasikan model tersebut memiliki gejala Multikolinieritas.

- Partial Correlation

  Cara kedua adalah dengan melihat keeratan hubungan antara dua variabel penjelas atau yang lebih dikenal dengan istilah korelasi.

2.11.4 Uji Heteroskedastisitas

Untuk uji Heteroskedastisitas, seperti halnya uji normalitas, cara yang sering degunakan dalam menentukan apakah suatu model terbebas dari masalah heteroskedastisitas atau tidak hanya dengan melihat Scatter Plot dan dilihat apakah residual memiliki pola tertentu atau tidak. Cara ini menjadi fatal karena pengambilan keputusan apakah suatu model terbebas dari masalah heteroskedastisitas atau tidak hanya berpatok pada pengamatan gambar saja tidak dapat dipertanggungjawabkan kebenarannya. Banyak metode statistik yang dapat digunakan untuk menentukan apakah suatu model terbebas dari masalah heteroskadastisitas atau tidak, seperti misalnya Uji White, Uji Park, Uji Glejser, dan lain-lain.

Salah satu uji heteroskedastisitas yang mudah dan dapat digunakan dengan SPSS, adalah Uji Glejser. Uji Glejser secara umum dinotasikan sebagai berikut:

$$|e| = b_1 + b_2x_2 + v$$

dimana:

$|e|$ = nilai Absolut dari rsidual yang dihasilkan dari regresi model

x_2 = variabel penjelas

Bila variabel penjelas secara statistik signifikan mempengaruhi residual maka dapat dipastikan model ini memiliki masalah Heteroskedasitas.

BAB 5

PENUTUP

5.1 Simpulan

Berdasarkan hasil pembahasan pada Bab 4, diperoleh simpulan bahwa berdasarkan metode yang dipilih untuk digunakan yaitu metode *Maximum Likelihood Estimation* (MLE) untuk data tersensor tipe II untuk data berdistribusi *Weibull* dengan ukuran sampel penelirian (n) individu, t tahun, dan r sensor berupa $t_{\mathrm{r}} = 90$ maka diambil kesimpulan sebagai berikut:

1. Metode estimasi menggunakan *Maximum Likelihood Estimation* (MLE) dapat digunakan untuk melakukan pendugaan parameter dengan baik bergantung pada fungsi kepadatan peluang, fungsi *hazard* dan fungsi *survival* dari distribusi data.
2. Estimasi parameter data tersensor tipe II berdistribusi Weibull menggunakan metode *Maximum Likelihood Estimation* (MLE) untuk data tersensor tipe II pada data historis tabel mortalitas CSO 1941 menghasilkan nilai estimasi $\hat{\alpha} = -1{,}57 * 10^{-11}$ dan $\hat{\beta} = 5{,}4$.

5.2 Saran

1. Perlu dikembangkan fungsi *survival* dari distribusi selain Weibull yang lebih representatif dalam menggambarkan perilaku data.
2. Mengembangkan metode *Maximum Likelihood Estimation* (MLE) data tersensor tipe II untuk distribusi selain Weibull dalam data uji hidup.

DAFTAR PUSTAKA

Aunon, J. I.,& Chandrasekar, V. 1997. *Introduction to Probability and Random Processes*. McGraw-Hill Companies, Inc.

Asadi, S. 2011. Estimation of the Weibull Distribution Based on Type-II Censored Samples. *Journal of Applied Mathematical Sciences*, 5(52):2549-2558.

Baim, L.J.,& Engelhard, M. 1991. *Introduction to Probability Mathematical Statistics*. 2nd ed. Dux Bury Press. Belmont, California.

Brown, R.L. 1997. *Introduction to the Mathematics of Demography*. 2nd ed. Winsted: Actec Publication.

Coale, A.J., and Paul, D. 1983. *Regional Model Life Tables and Stable Population*. 2nd ed. New York: Academic Press.

Edward, J. Dudewicz., and Satya, Mishra. 1988. *Modern mathematical statistik*. Wiley: The University of Michigan Press.

Lawless, J.F. 1982. *Statistical Model and menthods for lifetime data*. New York: John Wiley and Sons, inc.

Lawless, J.F. 2003. *Statistical Model and menthods for lifetime data*. 2nd edition. New York: John Wiley and Sons, inc.

Muller, H.G. 2004. *Demographic Window to aging in the wild: Contructing Life Tables and Estimating Survival Functions from Marked Individuals of Know Age*. *J Aging Cell*. Vol 3:125-131.

Noveria, Mita. 2004.*Dasar-dasar Demografi dan Dinamika mobilitas penduduk di wilayah perbatasan*. LIPI.

Razali, A.M., Salih, A.A., and Mahdi A.A. 2009. Estimation Accuracy of Weibull Distribution Parameter. *Journal of Applied Sciences Research*, 5(7):790-795.

Santoso, Singgih. 2000. *Buku Latihan SPSS Statistik PArametrik*. Jakarta: PT. Elex Media Kamputindo.

Scot & Nowak. 2004. *Introduction to Bayesian Statistic and Estimation*. 2th ed. New York: Pearson Education International.

Turmudi dan Harini, S. 2008. *Metode Statistik pendekatan Teoritis dan Alikatif*. Malang: UIN Malang press.

Wallpole, R.E.,& Myers, R.H. 2007. *Probability and Statistics for Engineers and Scientist*. 8th ed. Pearson Education International.

William, W & Dauglas, C. 1990. *Probabilita dan Statistik dalam ilmu rekayasa dan manajemen*. 2nd edition. Universitas Indonesia.