BAB II
TEORI PENUNJANG

2.1 Variabel Acak (random variable)

Suatu hasil pengumpulan data berupa suatu bilangan, baik merupakan kegiatan riset maupun eksperimen. Suatu contoh dalam pelemparan koin ditunjukkan beberapa bilangan riil x untuk setiap elemen dalam ruang sampel. Pada umumnya, bilangan riil x menunjukkan setiap hasil,e dari ruang sampel ζ. Suatu notasi fungsi bilangan riil x adalah: x = X(e), X adalah fungsi dari hasil e. Daerah asal X adalah ζ dan bilangan-bilngan dalam rangenya adalah bilangan-bilangan riil. Fungsi X diatas disebut variabel random. Variabel random terdiri dari variabel random diskrit maupun variabel random kontinu. Jika ruang range R_X,Variabel random X adalah terbatas atau tidak terbatas yang dapat dihitung, maka X disebut Variabel Random diskrit. Jika ruang range R_X , variabel random X adalah sebuah interval atau kumpulan dari interval-interval, maka X disebut variabel random kontinu.

Definisi 2.1.1

Jika ε sebuah percobaan yang memiliki ruang sample ζ dan x sebuah fungsi yang dinotasikan sebuah bilangan riil X(e) untuk setiap hasil e $\in \zeta$, maka X(e) disebut sebuah variabel random.

Definisi 2.1.2

Jika X adalah sebuah variabel random diskrit yang dihubungkan dengan sebuah bilangan $p_x(x_I) = P (X = x_i)$ dengan masing-masing hasil x_i, dalam R_x untuk i = 1,2,……, m, …. Dimana bilangan $p_x(x_i)$ memenuhi:

1. $p_x(x_I) \geq 0$ Untuk seluruh i.
2. $\sum_{i=1}^{x} p_x(x_i) = 1$

Definisi 2.1.2.a

Untuk sebuah variabel random kontinu X, didefinisikan:

$$P(a \leq X \leq b) = \int_a^b fx(x)dx$$

Dimana fungsi fx dinyatakan sebagai fungsi densitas probabilitas (fdp), yang memenuhi kondisi-kondisi sebagai berikut:

1. $f_x(x) \geq 0$ untuk setiap $x \in R_x$
2. $\int_{R_x} f_x(x)\,dx = 1$

2.2 Ekspekstasi (Nilai Harapan)

Jika X adalah sebuah variabel random, dan Y = H(x) adalah sebuah fungsi dari X, maka nilai harapan (expected value) H(x) didefinisikan sebagai berikut:

$$E[H(x)] = \sum_{\text{seluruh } i} H(x_i).p(x_i), \qquad \text{untuk X yang diskrit.}$$

$$E[H(x)] = \int_{-\infty}^{\infty} H(x)f(x)dx, \qquad \text{untuk X yang kontinu.}$$

Jika $H(X) = X$, maka $E[H(x)] = E(x) = \mu$, sehingga nilai harapan variabel random X adalah rata-rata μ.

Jika $H(x) = (x - \mu)^2$, maka $E[(x - \mu)^2] = \sigma^2$.

Jika $H(x) = x$, maka $var(x) = E[(x-E(x))^2] = E(x^2) - [E(x)]^2$

2.3 Penarikan Sampel

Pengumpulan data dengan cara sensus, walaupun dapat diperoleh data yang sebenarnya, biasanya banyak mengeluarkan biaya, waktu dan tenaga, sehingga dalam prakteknya sering digunakan penarikan sampel yang akan memberikan nilai taksiran atau penduga. Metode penarikan sampel lebih praktis, lebih mudah dan memerlukan tenaga yang lebih sedikit dibandingkan dengan sensus.

Suatu himpunan bagian pengamatan dipilih dari sebuah populasi disebut dengan sampel. Data penarikan sampel (sampling) merupakan nilai penduga karena adanya kesalahan penarikan sampel (sampling error).

Penarikan sampel acak (sampel random) sederhana adalah penarikan sampel dimana pemilihan elemen-elemen populasi dilakukan sedemikian rupa sehingga setiap elemen mempunyai peluang yang sama untuk dipilih.

2.4 Estimasi Parameter

Estimasi parameter merupakan salah satu bagian besar dari inferensi statistik. Inferensi Statistik merupakan proses memperoleh informasi dari data sampel yang digunakan untuk menarik kesimpulan tentang populasi dari sampel yang dipilih.

Suatu perkiraan tunggal pada sebuah parameter populasi adalah nilai tunggal numerik pada sebuah statistik yang berhubungan dengan parameter tersebut. Perkiraan tunggal adalah sebuah pemilihan yang dipilih untuk sebuah nilai parameter populasi yang tidak diketahui. Lebih jelasnya, jika X sebuah variabel random dengan sebaran peluang f(x), mempunyai parameter θ yang tidak diketahui, dan jika $X_1, X_2, \ldots\ldots, X_m$ sebuah sampel acak yang besarnya m dari X, maka statistik $\theta = h(X_1, X_2, \ldots\ldots, X_m)$ yang berhubungan dengan θ disebut estimator θ. Perhatikan bahwa estimator θ adalah sebuah peubah acak, karena estimator tersebut merupakan sebuah fungsi data sampel. Setelah sampel dipilih, θ diperoleh berdasarkan nilai tertentu yang disebut perkiraan tunggal θ.

Sebagai contoh, misalkan sebuah variabel random X yang memiliki distribusi normal dengan rataan μ yang tidak diketahui dan varian σ^2 diketahui. Rataan sampel $\overline{X}$ adalah suatu estimator tunggal dari rataan populasi μ yang tidak diketahui, maka $\mu = \overline{X}$. Sesudah sampel dipilih, nilai numerik x adalah perkiraan tunggal μ. Demikian juga varian populasi σ^2 diperoleh dari varian sampel s^2, dan nilai numerik s^2 dihitung berdasarkan data sampel yang merupakan perkiraan tunggal σ^2.

Salah satu sifat yang diinginkan estimator yaitu menjadi tertutup, dalam beberapa pengertian pada nilai sebenarnya dari parameter yang tidak diketahui. Singkatnya dapat dikatakan bahwa θ merupakan sebuah estimator yang *unbias* dari parameter θ jika;

$$E(\theta) = \theta \qquad (2.1)$$

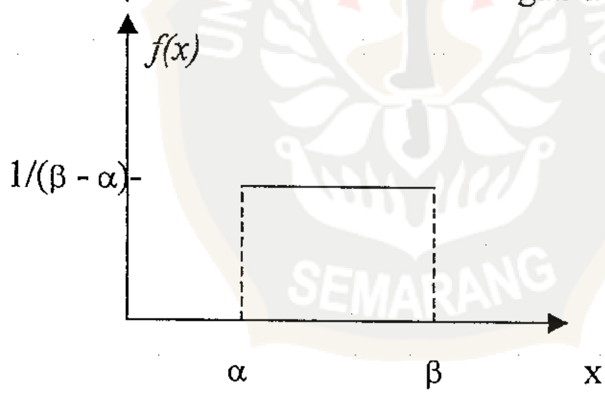
yaitu, θ merupakan sebuah estimator yang unbias dari θ, jika "pada rataan" nilainya adalah sama dengan θ, Dengan kata lain, rataan sebaran sampling θ sama dengan θ.

2.5 Distribusi Uniform

Distribusi uniform merupakan salah satu distribusi probabilitas kontinu. Distribusi uniform mempunyai daerah atau *range* dari suatu variabel kontinu X yang terdiri dari sebuah interval atau sebuah himpunan interval-interval. Fungsi densitas uniform didefinisikan sebagai;

$$f(x) = 1 / (\beta - \alpha) \qquad \alpha \leq x \leq \beta$$
$$= 0 \qquad \text{lainnya}$$

dimana α dan β adalah konstanta riil dengan $\alpha < \beta$.



Gambar sebuah densitas uniform.

Sebuah variabel random berdistribusi uniform mempunyai fungsi densitas probabilitas yang konstan sepanjang interval. Dari definisi, konstanta harus timbal balik dari panjang interval agar memenuhi syarat bahwa:

$$\int f(x)dx = 1$$

Sebuah variabel random yang berdistribusi uniform menunjukkan kontinu yang analog dengan hasil-hasil yang mempunyai kemungkinan yang sama dalam pengertian bahwa untuk setiap interval [a,b], dimana $\alpha \leq a < b \leq \beta$, $P(a \leq X \leq b)$ adalah sama untuk seluruh sub interval yang sama panjangnya.

$$P(a \leq X \leq b) = \int_a^b \frac{dx}{(\beta - \alpha)} = \frac{(b - a)}{(\beta - \alpha)}$$

Pernyataan bahwa *dipilih sebuah titik secara random pada* $[\alpha,\beta]$ mempunyai arti bahwa nilai yang dipilih, misal Y, berdistribusi uniform pada $[\alpha,\beta]$.

2.6 Pengertian Dasar Statistik Bootstrap

Metode bootstrap bertitik tolak atas dasar analogi antara sampel dan populasi darimana sampel tersebut diambil. Parameter yang digunakan dalam metode bootstrap ini meliputi pendugaan secara empirik semua distribusi sampling dari $\hat{\theta}$. Misal diberikan suatu data observasi sebagai berikut : $x_1, x_2, \ldots, x_m$ merupakan suatu variabel independen dengan distribusi F(X). Suatu parameter $\hat{\theta} = g(F(.))$ digunakan untuk mengestimasi suatu parameter $\hat{\theta} = \hat{\theta}\,(x_1, x_2, \ldots, x_m)$. Ilustrasinya adalah sebagai berikut:

DUNIA NYATA			DUNIA BOOTSTRAP	
Sebaran Peluang Tidak diketahui	Pengamatan Sampel Acak		Sebaran Empirik	Sampel Bootstrap
$F \rightarrow$	$x = (x_1, \ldots\ldots, x_n)$	$\rightarrow$	$\hat{F} \rightarrow$	$x^* = (x^*_1, \ldots\ldots, x^*_n)$
	$\hat{\theta} = s(x)$			$\hat{\theta}^* = s(x^*)$
	Statistik yang diinginkan			Ulangan Bootstrap

Pengambilan sampel ulang (resample) berukuran m dilakukan dengan pengembalian beberapa kali, sehingga terdapat kemungkinan terdapat suatu data sampel terambil kembali. Banyaknya data yang dilakukan dalam satu pengambilan akan menghasilkan data baru sejumlah data observasi, dengan fungsi empiris $F_m(X)$ dan dinotasikan sebagai sampel bootstrap sebagai berikut : $x^*_1, \ldots\ldots, x^*_m$ dimana x^*_i berdistribusi $F_m(X)$.

Metode bootstrap yang ideal adalah dengan memperlakukan semua kemungkinan sampel sebanyak m^m yang ada. Namun dapat dibayangkan, misalnya terdapat m=30 maka akan terdapat 30^{30} kemungkinan sampel. Sehingga memakan waktu pengerjaan yang lama. Ini menjadi tidak efektif lagi, sehingga akhirnya digunakan pendekatan simulasi Monte Carlo, dimana semua kemungkinan sampel bootstrap menjadi sejumlah B yang cukup besar tetapi cukup kecil jika dibandingkan dengan jumlah bootstrap ideal.

2.7 Alasan Penggunaan Piranti Lunak MathCAD dan Excel

Pemakaian komputer untuk menyelesaian model matematika berupa persamaan regresi untuk memperoleh suatu analisa statistika, memerlukan suatu pendekatan dengan matriks terhadap vektor data pengamatan sehingga berhubungan dengan array memerlukan prosedur penyelesaian yang panjang menjadi lebih singkat, efisien dan efektif. Komputasi statistik memerlukan suatu input data dari suatu piranti lunak spreedsheet dalam penyajian data pengamatan. Piranti lunak spreedsheet yang dipakai MS Excel. Piranti ini dipilih karena mempunyai fasilitas penyimpanan file berupa kumpulan record data dalam bentuk tabel. Proses penyimpanan file data dengan fasilitas fixed delimited menghasilkan file berbentuk ***nama_file.PRN***.

Salah satu piranti lunak yang digunakan dalam komputasi statistik adalah MathCAD. Aplikasi ini memberikan suatu bentuk penyelesaian dalam menangani rumus-rumus, teks dan plot gambar. Dengan cara mendefinisikan variabel dan menghitung hasilnya, MathCAD dapat memecahkan problem-problem numerik dengan mudah.

Seperti piranti lunak lainnya, MathCAD mempunyai aturan penulisan dalam penyusunan penyelesaian model matematika. Salah satu kemudahannya adalah membuat proses penyelesaian dalam lembar kerja MathCAD berupa model matematika dengan mendefinisikan persamaan tersebut dan menampilkan hasilnya dengan mengetikkan variabel definisi diikuti dengan mengetik tanda "=". Misal untuk menghitung koefisien regresi sederhana dengan suatu pemanggilan data dari luar aplikasi MathCAD, kegiatannya adalah sebagai berikut;

Keyboard	Hasil di layar
data[shift][;]READPRN(DATA)	**data:=READPRN(DATA)**
X[shift][;]dataX	**X:=dataX**
Y[shift[;]dataY	**Y:=dataY**
β [shift][;](X^T*X)^-1 X^TY	**β :=$(X^TX)^{-1}X^TY$**

Untuk melihat hasilnya cukup dengan mengetik "β =". Dibandingkan dengan menggunakan bahasa program seperti BASIC, PASCAL atau bahasa pemrograman yang lain, model persamaan diatas memakan banyak baris program.