BAB II

TEORI PENUNJANG

Pada bab ini akan disajikan beberapa konsep dasar statistik yang mendukung untuk pembahasan selanjutnya mengenai standar error.

2.1 Pengertian Standar Error

2.1.1 Standar Error dari Mean

Standar error merupakan standar deviasi dari statistik. Sehingga yang perlu diingat bahwa standar error ini adalah sejenis standar deviasi yang mempunyai tugas yang sama seperti standar deviasi. Standar error merupakan suatu cara untuk mengindikasikan ketelitian statistik.

Jika X adalah sebuah variabel random dengan distribusi probabilitas F, dan Y = H(X) adalah sebuah fungsi dari X, maka nilai harapan atau ekspektasi H(X) didefinisikan sebagai berikut :

$$E_F[H(X)] = \sum_{seluruh\, i} H(x_i).P_F(x_i) \qquad \text{untuk X yang diskrit} \qquad (2.1.1.1)$$

$$E_F[H(X)] = \int_{-\infty}^{\infty} H(x).P_F(x) \qquad \text{untuk X yang kontinu} \qquad (2.1.1.2)$$

Karena yang akan dibahas selanjutnya mengenai distribusi kontinu, maka yang akan dijelaskan hanya untuk kasus X yang kontinu. Kemudian dibentuk H sehingga Y = H(X) adalah sebuah variabel random yang kontinu.

Jika H(X) = X, maka didapat :

$$E_F[H(X)] = E_F(X) = \mu_F \quad (2.1.1.3)$$

Oleh karena itu, nilai harapan atau nilai ekspektasi variabel random X adalah rata-rata μ_F. Jika $H(X) = (X-\mu_F)^2$, maka :

$$E[H(X)] = E[(X - \mu_F)^2] = \sigma_F^2 \quad (2.1.1.4)$$

Maka varian dari variabel random X dapat didefinisikan dalam susunan dari ekspektasi. Sejak varian digunakan secara luas, hal ini biasanya untuk memperkenalkan sebuah varian operator V yang didefinisikan dalam susunan nilai harapan atau ekspektasi operator E :

$$V[H(X)] = E\{[H(X) - E(H(X))]^2\} \quad (2.1.1.5)$$

Jika H(X) = X, maka :

$$V(X) = E[(X - E(X))^2]$$

$$= E(X^2) - [E(X)]^2 \quad (2.1.1.6)$$

yang merupakan varian X, dan dinotasikan dengan σ^2.

Jika $(x_1,x_2,\ldots,x_n)$ adalah sebuah sampel random yang berukuran n dari distribusi probabilitas F, maka mean dari sampel adalah :

$$\bar{x} = \sum_{i=1}^{n} x_i / n \quad (2.1.1.7)$$

Jika distribusi dari mean diatas diasumsikan normal dengan n besar, maka :

$$M\sum_{i=1}^{n}\frac{x_i}{n}(t)=M\sum_{i=1}^{n}x_i(\frac{t}{n})$$

$$=\prod_{i=1}^{n}Mx_i(\frac{t}{n})$$

$$Mx(t)=e^{\mu_F t+\frac{1}{2}\sigma_F^2 t^2}$$

$$M\sum_{i=1}^{n}\frac{x_i}{n}(t)=e^{\mu_F \frac{t}{n}+\frac{1}{2}\sigma_F{}^2(\frac{t}{n})^2}.e^{\mu_F \frac{t}{n}+\frac{1}{2}\sigma_F{}^2(\frac{t}{n})^2}\dots e^{\mu_F \frac{t}{n}+\frac{1}{2}\sigma_F{}^2(\frac{t}{n})^2}$$

$$=e^{\mu_F t+\frac{1}{2}\sigma_F{}^2 n(\frac{t}{n})^2}$$

$$=e^{\mu_F t+\frac{1}{2}\frac{\sigma_F{}^2}{n}t^2}$$

sehingga didapat nilai ekspektasi adalah μ_F dan variansinya $\frac{\sigma_F^2}{n}$.

Dengan kata lain ekspektasi dari $\bar{x}$ adalah sama seperti ekspektasi sebuah x, tetapi variansi dari $\bar{x}$ adalah 1/n kali variansi x.

Sedangkan standar error dari mean yaitu $\bar{x}$ adalah $se_F(\bar{x})$ atau $se(\bar{x})$ yaitu akar pangkat dua variansi $\bar{x}$,

$$se_F(\bar{x})=[\text{var}_F(\bar{x})]^{1/2}=\sigma_F/\sqrt{n} \qquad (2.1.1.8)$$

2.1.2 Standar Error dalam Teori Pengambilan Sampel

Pengertian standar error merupakan suatu pengertian yang memegang peranan yang begitu penting didalam teori pengambilan sampel. Di dalam membicarakan mengenai standar error itu, akan dibedakan dua macam sampel, yaitu sampel berukuran besar dan sampel berukuran kecil. Sedangkan distribusi sampel dibedakan menjadi dua macam, yaitu distribusi sampel eksperimental dan teoritis.

Distribusi sampel eksperimental adalah distribusi mengenai statistik sampel-sampel yang ditarik dari sebuah populasi. Perkataan 'eksperimental' digunakan karena penyusunan distribusi itu didasarkan atas eksperimen, bukan atas pertimbangan teoritis. Distribusi sampel eksperimental dapat dibentuk untuk setiap macam titik sampel. Dari distribusi sampel eksperimental ini dapat dihitung harga rata-rata hitung dan standar deviasi, yaitu :

$$\bar{x} = \sum_{i=1}^{n} x_i (f_i / n) \tag{2.1.2.1}$$

$$s_{\bar{x}} = \sqrt{\sum_{i=1}^{n} x_i^2 (f_i / n) - \bar{x}^2} \tag{2.1.2.2}$$

dimana $\bar{x}$ adalah harga rata-rata hitung dan f_i/n merupakan probabilitasnya yaitu frekuensi dibagi banyaknya data, serta $s_{\bar{x}}$ adalah standar deviasinya. Standar deviasi dari statistik ini biasanya dinamakan 'standar error' dari statistik itu. Harus diingat bahwa standar error ini adalah sejenis standar deviasi yang tentu mempunyai tugas yang sama seperti standar deviasi.

Bentuk lain dari distribusi sampel adalah distribusi sampel teoritis, yaitu sebuah distribusi yang hampir serupa dengan distribusi sampel eksperimental, hanya berbeda karena penggantian kolom frekuensi (atau kolom frekuensi relatif) dengan kolom probabilitas. Pembentukan kolom probabiltas itu didasarkan atas pertimbangan teoritis, bukan secara eksperimental. Didalam menyelidiki distribusi sampel, sebaiknya distribusi sampel teoritis ini yang dipakai. Akan tetapi karena untuk beberapa hal, distribusi seperti itu tidak dapat atau terlalu sulit dibentuk, maka sering terpaksa memakai distribusi sampel eksperimental, yang dapat dibentuk dengan melakukan eksperimen berulang-ulang. Distribusi sampel eksperimental itu hanyalah pengganti bagi distribusi teoritis yang mana distribusi yang terakhir ini tidak dapat dibentuk atau terlalu sulit dibentuk.

Misalkan populasi berukuran terhingga N dengan rata-rata μ dan simpangan baku σ. Dari populasi ini diambil sampel random berukuran n. Jika sampling dilakukan tanpa pengembalian, maka ada $\binom{N}{n}$ buah sampel yang berlainan. Untuk semua sampel yang didapat, masing-masing dihitung rata-ratanya. Dengan demikian diperoleh $\binom{N}{n}$ buah rata-rata. Anggap semua rata-rata ini sebagai data baru, jadi didapat kumpulan data yang terdiri atas rata-rata dari sampel-sampel. Dari kumpulan ini dihituing rata-rata dan simpangan bakunya. Jadi didapat rata-rata dari rata-rata, yang diberi simbol $\mu_{\bar{x}}$ dan simpangan baku dari rata-rata, yang diberi simbol $\sigma_{\bar{x}}$.

Sebagai contoh jika diberikan sebuah populasi dengan N = 10 dengan data sebagai berikut : 98, 99, 97, 98, 99, 98, 97, 97, 98, 99. Jika dihitung populasi ini mempunyai $\mu = 98$ dan $\sigma = 0{,}78$. Diambil sampel berukuran n = 2. Semuanya ada

* standar deviasi = akar kuadrat + dari varians
positif

$$\sigma = \sqrt{\frac{\Sigma (x-\mu)^2}{N}}$$

$\binom{10}{2} = 45$ buah sampel. Untuk setiap sampel dihitung rata-ratanya. Data dalam tiap sampel dan rata-rata tiap sampel diberikan dalam daftar berikut ini :

SAMPEL	RATA2	SAMPEL	RATA2	SAMPEL	RATA2
(98,99)	98 ½	(99,98)	98 ½	(99,98)	98 ½
(98,97)	97 ½	(99,99)	99	(99,97)	98
(98,98)	98	(97,98)	97 ½	(99,97)	98
(98,99)	98 ½	(97,99)	98	(99,98)	98 ½
(98,98)	98	(97,98)	97 ½	(99,99)	99
(98,97)	97 ½	(97,97)	97	(98,97)	97 ½
(98,97)	97 ½	(97,97)	97	(98,97)	97 ½
(98,98)	98	(97,98)	97 ½	(98,98)	98
(98,99)	98 ½	(97,99)	98	(98,99)	98 ½
(99,97)	98	(98,99)	98 ½	(97,97)	97
(99,98)	98 ½	(98,98)	98	(97,98)	97 ½
(99,99)	99	(98,97)	97 ½	(97,99)	98
(99,98)	98 ½	(98,97)	97 ½	(97,98)	97 ½
(99,97)	98	(98,98)	98	(97,99)	98
(99,97)	98	(98,99)	98 ½	(98,99)	98 ½
Jumlah semua rata-rata = 4410					

Jumlah ke-45 buah rata-rata = 4.410 . Maka rata-ratanya untuk ke-45 rata-rata ini adalah :

$$\frac{4410}{45} = 98$$

Jadi $\mu_{\bar{x}} = 98$.

Simpangan baku yang didapat adalah :

$$\sigma_{\bar{x}} = 0{,}52$$

Tetapi rata-rata populasi $\mu = 98$ dan simpangan baku $\sigma = 0{,}78$. Selanjutnya dihitung:

$$\frac{\sigma}{n}\sqrt{\frac{N-n}{N-1}} = \frac{0{,}78}{\sqrt{2}}\sqrt{\frac{10-2}{10-1}} = 0{,}52$$

Dari contoh soal di atas dapat diambil kesimpulan bahwa :

$$\mu_{\bar{x}} = \mu$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

yang merupakan hubungan tetap dan berlaku umum antara standar deviasi populasi dengan standar error dari nilai rata-rata hitung sampel. Rumus diatas berlaku jika sampel random berukuran n diambil dari sebuah populasi berukuran N dengan rata-rata μ dan simpangan baku (standar deviasi populasi) adalah σ, serta σ_x adalah standar error dari nilai rata-rata hitung sampel, jika n/N > 5 %.

Jika N cukup besar dibandingkan terhadap n atau $N \to \infty$, maka berlaku :

$$\mu_{\bar{x}} = \mu$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Rumus diatas sebaiknya digunakan bila n/N $\leq$ 5 %, yang merupakan hubungan yang berlaku umum antara nilai rata-rata hitung sampel dengan populasi.

Distribusi sampel teoritis dari x, untuk nilai-nilai n yang besar, dapat dihampiri dengan baik oleh sebuah distribusi normal ('Central Limit Theorem'). Hal ini hanyalah benar untuk sampel berukuran besar saja, seperti yang disebutkan oleh teorema itu :

"Central Limit Theorem" :

Jika n merupakan sebuah bilangan besar (dimana n menunjukkan besarnya sampel), maka distribusi sampel teoritis dari nilai rata-rata hitung $\bar{x}$ dari sampel itu dapat dihampiri dengan baik oleh sebuah distribusi normal.

Jika sampel yang dipakai terlalu kecil, tidak dapat dipakai distribusi normal untuk menghampiri distribusi sampel teoritis itu, tetapi harus memakai distribusi-t. Ini disebabkan standar error suatu statistik tidaklah bergantung hanya kepada jenis sampel yang dipakai saja, tetapi bergantung kepada besarnya sampel yang dipakai.

Distribusi-t itu, untuk sampel-sampel yang kecil, sangat mendekati distribusi normal. Untuk sampel berukuran 30 pengamatan atau lebih, distribusi-t itu sudah sedemikian dekatnya kepada distribusi normal, sehingga pemakaian kurve normal sebagai penggantinya sudah dapat dipertanggungjawabkan.

Biasanya didalam membicarakan teori pengambilan sampel, orang lebih sering memakai distribusi normal sebagai penghampir bagi distribusi sampel teoritis, atau menganggap bahwa populasi yang bersangkutan berdistribusi secara normal. Ada beberapa alasan yang dapat diajukan untuk itu, diantaranya :

(i) Distribusi normal berlaku agak umum, apalagi jika sampel yang dipakai cukup besar

(ii) Distribusi-distribusi lainnya dapat dihampiri dengan baik oleh distribusi normal.

(iii) Distribusi normal lebih mudah dipakai dibandingkan dengan distribusi-distribusi lainnya.

2.2 Estimasi Parameter

Estimasi parameter merupakan salah satu bagian besar dari teknik statisik induktif. Statistik induktif merupakan proses memperoleh informasi dari data sampel yang digunakan untuk menarik kesimpulan tentang populasi dari sampel yang dipilih. Dalam hal ini yang akan dijelaskan mengenai estimasi tunggal dan estimasi interval keyakinan serta estimasi standar error dai mean.

2.2.1 Estimasi Tunggal

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 unik untuk sebuah parameter populasi yang tidak diketahui. Lebih jelasnya jika x sebuah variabel random dengan distribusi probabilita f(x), mempunyai parameter θ yang tidak diketahui, dan jika $x_1, x_2, \ldots x_n$ sebuah sampel random yang besarnya n dari x, maka statistik $\hat{\theta} = h(x_1, x_2, \ldots x_n)$ yang berhubungan dengan θ disebut estimator θ. Perhatikan bahwa estimator θ adalah sebuah variabel (peubah) random, karena estimator tersebut merupakan sebuah fungsi data sampel. Setelah sampel dipilih, $\hat{\theta}$ diperoleh berdasarkan nilai tertentu yang disebut perkiraan tunggal θ.

Sebagai contoh, misalkan sebuah variabel random x yang memiliki distribusi normal dengan rata-rata $\hat{\mu}$ yang tidak diketahui dan varian τ^2 diketahui. Rata-rata sampel $\overline{X}$ adalah suatu estimator tunggal dari rata-rata populasi μ yang tidak diketahui, maka $\hat{\mu} = \overline{X}$. Sesudah sampel dipilih, nilai numerik $\overline{X}$ adalah perkiraan tunggal μ. Demikian juga jika varian populasi τ^2 tidak diketahui, sebuah estimator tunggal untuik τ^2 adalah varian sampel S^2, dan nilai numerik s^2 dihitung berdasarkan data sampel yang merupakan perkiraan tunggal τ^2. Jika rata-rata sebuah variabel random akan diestimasi, harus diperhatikan salah satu dari rata-rata sampel, median sampel atau mungkin observasi yang paling kecil dalam sampel sebagai estimator tunggal.

Sebuah sifat yang diinginkan estimator yaitu menjadi "tertutup" dalam beberapa pengertian pada nilai sebenarnya dari parameter yang tidak diketahui. Singkatnya dapat dikatakan bahwa $\hat{\theta}$ merupakan sebuah estimator yang unbias dari parameter θ jika :

$$E(\hat{\theta}) = \theta \qquad (2.2.1.1)$$

yaitu $\hat{\theta}$ merupakan sebuah estimator yang unbias dari θ, jika "pada rata-rata" nilainya adalah sama untuk θ. Dengan kata lain, rata-rata distribusi sampling $\hat{\theta}$ sama dengan θ.

Sebagai contoh, jika X adalah variabel random dengan rata-rata μ dan varian σ^2. Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel random yang besarnya n dari X. Dapat diperlihatkan bahwa rata-rata sampel X dan varian sampel S^2 adalah estimator yang unbias dari μ dan σ^2, yaitu bahwa :

$$E(\overline{X}) = \mu$$

dan $E(S^2) = \sigma^2$

Maka varian sampel S^2 adalah sebuah estimator yang *unbias* terhadap varian populasi σ^2. Tetapi standar deviasi sampel S adalah yang bias terhadap standar deviasi populasi σ. Untuk sampel yang besar bias ini dapat diabaikan.

2.2.2 Estimasi Interval Keyakinan

Dalam beberapa keadaan, sebuah perkiraan tunggal tidak memberikan cukup penjelasan tentang parameter yang menarik. Sebagai contoh dalam memperkirakan kekuatan beton, sebuah bilangan tunggal mungkin tidak mempunyai arti. Sebuah perkiraan interval dengan bentuk $L \leq \mu \leq U$ mungkin lebih berguna.

Pada umumnya untuk mengerjakan suatu estimator interval parameter tidak diketahui θ, harus ditentukan dua statistik L dan U sebagai berikut :

$$P\{L \leq \theta \leq U\} = 1 - \alpha \qquad (2.2.2.1)$$

Interval yang dihasilkan adalah :

$$L \leq \theta \leq U \qquad (2.2.2.2)$$

disebut interval keyakinan 100(1- α) persen untuk parameter θ yang tidak diketahui. L dan U disebut batas keyakinan atas dan bawah, dan 1-α disebut koefisien keyakinan. Interpretasi sebuah interval keyakinan yaitu jika beberapa sampel random dikumpulkan dan sebuah interval keyakinan 100(1-α) persen pada θ dihitung dari setiap sampel, maka 100(1-α) persen interval ini akan berisi nilai θ sebenarnya.

Panjang interval keyakinan yang diobservasi adalah sebuah ukuran penting kualitas informasi yang diperoleh dari sampel. Luas setengah interval θ-L atau U-θ disebut ketepatan estimator. Lebih panjang interval, lebih kecil informasi yang diperoleh mengenai θ yang sebenarnya. Pada suatu keadaan yang ideal, dapat diperoleh suatu interval yang pendek dengan tingkat keyakinan yang tinggi.

2.2.3 Estimasi Standar Error dari Mean

Seperti yang telah dijelaskan diatas bahwa standar error dari mean ($\bar{x}$) adalah $se_F(\bar{x}) = \sigma_F / \sqrt{n}$. Kemudian dengan menggunakan prinsip distribusi pengganti, disubstitusi $\hat{F}$ untuk F ke dalam formula $se_F(\bar{x}) = \sigma_F / \sqrt{n}$. Estimasi pengganti dari $\sigma_F = [E_F(x - \mu_F)^2]^{1/2}$ adalah

$$\hat{\sigma} = \sigma_{\hat{F}} = \{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\}^{1/2} \qquad (2.2.3.1)$$

Jika $\mu_{\hat{F}} = \bar{x}$ dan $E_{\hat{F}}g(x) = \frac{1}{n}\sum_{i=1}^{n} g(x_i)$ untuk fungsi g, maka estimasi standar error $\hat{se}(\bar{x}) = se_{\hat{F}}(\bar{x})$,

$$\hat{se}(\bar{x}) = \sigma_{\hat{F}} / \sqrt{n} = \{\sum_{i=1}^{n}(x_i - \bar{x})^2 / n^2\}^{1/2} \qquad (2.2.3.2)$$

Formula (2.2.3.2) agak sedikit berbeda dari standar error yang diestimasi biasa, karena σ_F yang digunakan adalah $\hat{\sigma} = \sigma_{\hat{F}} = \{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\}^{1/2}$ bukan $\bar{\sigma} = \{\sum_{i=1}^{n}(x_i - \bar{x})^2/(n-1)\}^{1/2}$. Hal ini dikarenakan penggunaan $\hat{\sigma}$ dalam mengestimasi σ_F adalah sebaik $\bar{\sigma}$, hanya saja pembagian dengan n-1 membuat $\bar{\sigma}^2$ unbias untuk σ_F^2. Dalam uraian diatas, prinsip pengganti telah digunakan dua kali yaitu estimasi ekspektasi μ_F dengan $\mu_{\hat{F}} = \bar{x}$ dan kemudian estimasi standar error $se_F(\bar{x})$ *dengan* $se_{\hat{F}}(\bar{x})$.