harsawara@gmail.com

phone: 081321005577

STATISTIKA

Pertemuan 11
Pendugaan Parameter

Assalamu'alaikum, Class ..

Bagaimana kabarnya kalian hari ini? Semoga Allah senantiasa limpahkan kita dengan kesehatan dan kebahagiaan .. sehingga kita dapat terhindar dari hal-hal yang menjadikan kita tidak sehat dan tidak bahagia .. aamiiin ...

Oh iya, karena masih dalam momentum Idul Fitri, saya menghaturkan permohonan maaf lahir dan bathin atas segala khilaf dan salah .. semoga amal sholeh dan kebajikan yang kita upayakan selama bulan Ramadhan, Allah terima ..

baiklah, pada kesempatan kali ini kita akan membicarakan mengenai *Pendugaan Parameter atau Estimasi* .. detail materi ini sebenarnya menjadi lingkup bahasan di statistika untuk NonDik, dan nyaris tidak berkaitan secara langsung dalam kajian statistika untuk penelitian pendidikan .. tapi tidak ada salahnya saya memperkenalkan konsep ini ke kalian ..

Sebagaimana yang telah kita bahas di pertemuan awal dulu, secara umum metode statistik dikategorikan kedalam dua kelompok yaitu statistik deskriptif (*descriptive statistics*) dan statistik inferensia (*statistical inference*) ..

Statistik deskriptif merupakan metode yang berhubungan dengan pengumpulan dan penggambaran satu set data untuk mendapatkan informasi yang berarti .. statistik deskriptif hanya memberikan informasi dari data yang dikumpulkan (*sample*) dan tidak dapat memberikan kesimpulan atau inferensia untuk kelompok data yang lebih besar (populasi) ..

Metode yang berhubungan dengan analisis dari sebuah sampel yang akan memberikan prediksi atau inferensia tentang seluruh set data atau populasi disebut dengan statistik inferensia .. statistik inferensia sendiri sejatinya dapat dibagi menjadi dua kelompok, yaitu pendugaan (estimation) dan uji hipotesis (hypotheses testing) .. di pertemuan sebelumnya saya sempat memperkenalkan pengujian hipotesis, termasuk didalamnya potensi kesalahan dalam pengujian hipotesis .. nanti kita akan melanjutkan bahasan berkenaan pengujian hipotesis ini di pertemuan berikutnya.. hari ini kita berfokus dulu pada pendugaan parameter ya ..

What is Pendugaan Parameter?

Sebuah nilai statistik yang dihitung dari sebuah sample n merupakan penduga titik (*point estimation*) dari suatu parameter populasi ..

Contoh: statistik rerata sampel ($\bar{X}$) merupakan penduga titik dari parameter rerata populasi μ .. demikian juga dengan statistik lainnya misalkan statistik simpangan baku s adalah penduga titik dari parameter simpangan baku populasi σ ..

Statistik yang digunakan untuk mendapatkan sebuah estimasi titik disebut penduga (estimator) .. nilai dari sebuah estimator dikatakan pendugaan (estimate) .. dengan kata lain, sebuah estimator akan memberikan sebuah nilai yang disebut dengan estimate ..

Sebuah estimator diharapkan dapat menduga nilai parameter populasi tanpa kesalahan (error) .. namun demikian jika terdapat dua atau lebih estimator, kita harus memutuskan estimator mana yang lebih baik .. dari konsep ini, muncullah istilah *unbiased estimator* ..

Unbiased estimator adalah ukuran keakuratan statistik yang digunakan untuk memperkirakan parameter populasi .. atau dengan definisi yang lebih sederhana, dapat dikatakan bahwa unbiased estimator adalah selisih antara nilai parameter sebenarnya dengan nilai penduganya .. akurat dalam pengertian ini dapat dipahami bahwa nilai tersebut tidak terlalu tinggi atau terlalu rendah .. jika suatu nilai estimasi bernilai terlalu tinggi (*overestimate*) atau terlalu rendah (*underestimate*) maka pendugaan yang seperti ini disebut 'bias' ..

Jika terdapat dua *unbiased estimator* dari parameter populasi yang sama, maka estimator yang distribusi samplingnya mempunyai varians terkecil adalah yang menjadi pilihan. *Unbiased estimator* yang mempunyai varians terkecil dikatakan estimator yang paling efisien (*the most efficient estimator*) ..

Walaupun demikian, adalah sangat sulit melakukan pendugaan titik terhadap parameter populasi secara tepat ..

trus, bagaimana antisipasinya?

Muncullah konsep pendugaan interval (*interval estimate*) .. sebuah pendugaan interval dari parameter populasi θ adalah sebuah interval dalam bentuk θ1 < θ < θ2, dengan θ1 adalah batas bawah interval dan θ2 adalah batas atasnya .. jikalau dituliskan dalam bentuk probabilitas, maka P(θ1 < θ < θ2) = 1 - α, untuk 0 < α < 1 ..

(1- α) ini kemudian lebih sering dikenal sebagai koefisien keyakinan (*confidence coefficient*) atau batas tingkat keyakinan (*degree confidence limits*) .. perumusan perhitungan Coefficient Interval (CI - Interval Kepercayaan) ini kira-kira dapat disajikan sebagai berikut:

CI untuk μ, dimana σ diketahui

Rata-rata sampel $\bar{x}$ akan digunakan sebagai sebuah estimasi titik untuk rata-rata populasi μ. $\bar{x}$ akan menjadi penduga yang sangat akurat untuk μ ketika n besar, karena $\sigma_{\bar{x}}^2 = \sigma^2 / n$. Jika sampel diseleksi dari sebuah populasi normal atau jika n cukup besar, kita dapat membentuk sebuah CI untuk μ berdasarkan dari distribusi sampling $\bar{x}$. Distribusi sampling $\bar{x}$ aproksimasi berdistribusi normal dengan rata-rata $\mu_{\bar{x}} = \mu$ dan standar deviasi $\sigma_{\bar{x}} = \sigma / \sqrt{n}$, sehingga

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$$

$$P\left(-z_{\alpha/2} < \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} < z_{\alpha/2}\right) = 1 - \alpha$$

$$P(-z_{\alpha/2}.\sigma / \sqrt{n} < \bar{x} - \mu < z_{\alpha/2}.\sigma / \sqrt{n}) = 1 - \alpha$$

$$P(-\bar{x} - z_{\alpha/2}.\sigma / \sqrt{n} < -\mu < -\bar{x} + z_{\alpha/2}.\sigma / \sqrt{n}) = 1 - \alpha$$

$$P(\bar{x} - z_{\alpha/2}.\sigma / \sqrt{n} < \mu < \bar{x} + z_{\alpha/2}.\sigma / \sqrt{n}) = 1 - \alpha$$

Jika $\bar{x}$ digunakan sebagai penduga μ, maka kita dapat (1- α)100% yakin bahwa kesalahan atau selisih antara μ dan $\bar{x}$ tidak melebihi $z_{\alpha/2}.\sigma / \sqrt{n}$. Untuk sampel kecil yang dipilih dari populasi tidak normal, derajat keyakinannya menjadi tidak akurat. Sedangkan untuk sampel besar (n ≥ 30) teori sampling menjamin hasil yang baik tanpa melihat sebaran populasinya. Untuk menghitung disebut (1- α)100% CI untuk μ, diasumsikan bahwa σ diketahui. Jika n ≥ 30, σ dapat diganti dengan standar deviasi sampel s.

ngerti gak? Kalo gak ngerti, nanti kita diskusikan lebih lanjut di pertemuan berikutnya ya ..

Misalkan, kita punya kasus diketahui rata-rata dan standar deviasi Indeks Prestasi Kumulatif (IPK) dari 36 mahasiswa yang diambil secara random adalah 2,6 dan 0,3 .. tentukan 95% dan 99% interval kepercayaan untuk rata-rata seluruh kelas!

Solusi: untuk 95% CI

$$P\left(2{,}6-(1{,}96).(0{,}3/\sqrt{36}) < \mu < 2{,}6+(1{,}96).(0{,}3/\sqrt{36}\right)=0{,}95$$

$$P(2{,}50 < \mu < 2{,}70)=0{,}95$$

untuk 99% CI

$$P\left(2{,}6-(2{,}575).(0{,}3/\sqrt{36}) < \mu < 2{,}6-(2{,}575).(0{,}3/\sqrt{36}\right)=0{,}99$$

$$P(2{,}47 < \mu < 2{,}73)=0{,}99$$

Disini dapat kita lihat bahwa interval yang lebih panjang diperlukan untuk menduga μ dengan tingkat keyakinan yang lebih tinggi. Dari contoh diatas dapat kita katakan 95% tingkat keyakinan bahwa rata-rata sampel $\bar{x}$ =2,6 berbeda dari μ tidak lebih dari 0,1 dan 99% tingkat keyakinan bahwa perbedaannya tidak lebih dari 0,13. Seringkali kita berharap untuk mengetahui seberapa besar sampel yang diperlukan agar kesalahan dalam menduga μ tidak melebihi besaran e yang dikehendaki. Hal ini sebenarnya sama dengan menentukan n dimana $z_{\alpha/2}.\sigma/\sqrt{n}=e$. Jika $\bar{x}$ digunakan sebagai penduga μ, maka kita dapat (1- α)100% yakin bahwa kesalahan tidak melebihi besaran e ketika jumlah sampel sebesar

$$n=\left(\frac{z_{\alpha/2}.\sigma}{e}\right)^2$$

Oh iya, yang harus kalian perhatikan adalah nilai dari Z(a/2) .. untuk (1- a) = 0.95, maka dengan mudah dapat kita ketahui bahwa nilai a adalah 5% atau 0.05 .. kalian hanya perlu mengoperasikan secara biasa saja untuk mendapatkan nilai a ini ..

Nah, nilai Z(a/2) diperoleh dari Z(0.05/2) = Z(0.025) .. silahkan kalian lihat di tabel Z, berapa nilai Z untuk luas daerah di bawah kurva 1 - 0.025 = 0.975 .. maka dengan mudah kalian akan memperoleh nilai Z-nya adalah 1.96 .. atau dengan kata lain nilai dari Z(a/2) untuk (1- a) = 0.95 adalah 1.96 ..

Cara yang sama juga berlaku untuk (1- a) = 0.99 .. Z(a/2) untuk (1- a) = 0.99 adalah 2.58 ..

Jika didasarkan pada formula tersebut, maka kita dapat menentukan nilai n jika varians populasi diketahui .. sementara jika varians populasi tidak diketahui maka kita dapat gunakan dugaan dari σ dari sebuah sampel terdahulu yang besarannya n ≥ 30.

Contoh:
Berapa besar sampel yang diperlukan pada kasus sebelumnya jika kita menginginkan 95% tingkat keyakinan bahwa dugaan dari μ tidak berbeda lebih dari 0,05?

Solusi: diketahui standar deviasi sampel s = 0,3 yang didapat dari sampel sebelumnya yang berjumlah 36.

$$n = \left(\frac{(1,96).(0,3)}{0,05} \right)^2 = 138,3$$

Dapat dikatakan bahwa kita dapat 95% yakin bahwa dengan sampel random sebesar 139 akan memberikan sebuah dugaan $\bar{x}$ yang berbeda tidak lebih dari 0,05 terhadap μ.

Gimana jika jumlah sampelnya relatif kecil, misalkan 12?

Untuk sampel yang berjumlah kecil, secara formulasi pendugaan parameter tidak berbeda dengan sampel berjumlah besar .. hanya saja tabel acuannya yang berubah .. jika sampel berjumlah besar, kita menggunakan tabel Z, maka untuk sampel berjumlah kecil kita menggunakan tabel T ..

Interval konfidensi inilah yang biasanya kita kenal dengan sebutan *margin of error* (MoE) atau batas toleransi kesalahan yang masih bisa diterima ..

Gimana?
mudah-mudah cukup mengambarkan dan menjelaskan .. jangan keburu mabok! :p

jikalau ada hal-hal yang belum dimengerti, kalian boleh menanyakannya langsung ke saya ..

tetap semangat dan stay safe & health ya .