INTERVAL KEPERCAYAAN

Untuk menaksir interval taksiran parameter θ dengan koefisien kepercayaan γ, maka sebuah sampel acak diambil, lalu hitung nilai-nilai statistik yang diperlukan. Perumusan dalam bentuk peluang untuk parameter θ antara A dan B:

$$P(A < \theta < B) = \gamma$$

dengan A dan B *fungsi dari statistik*, jadi merupakan *variabel acak*, tidak bergantung pada θ.

Arti dari formula di atas adalah secara $\gamma\%$ percaya bahwa parameter θ akan ada didalam interval (A, B). Jadi tidaklah dikatakan: peluangnya sama dengan γ bahwa θ terletak A dan B, melainkan seseorang hamya yakin $\gamma\%$ bahwa θ itu terletak antara A dan B.

1. Interval Kepercayaan Bagi Rata-Rata

Misalkan sebuah populasi berukuran N dengan rata-rata μ dan simpangan baku σ. Dari populasi ini parameter μ akan ditaksir. Untuk keperluan ini, diambil sampel acak berukuran n, lalu dihitung statistik yang perlu, ialah $\bar{x}$ dan s. Titik taksiran untuk rata-rata μ ialah $\bar{x}$. Dengan kata lain, nilai μ besarnya ditaksir oleh harga $\bar{x}$ yang didapat dari sampel.

Untuk memperoleh taksiran yang lebih tinggi derajat kepercayaan, digunakan interval taksiran atau selang taksiran disertai nilai koefisien kepercayaan yang dikehendaki. Dibedakan menjadi tiga hal

a. Simpangan baku σ diketahui dan populasinya berdistribusi normal

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

Dengan γ = koefisien kepercayaan dan $z_{\frac{1}{2}\gamma}$ = bilangan z didapat dari tabel normal baku untuk peluang $\frac{1}{2}\gamma$.

Untuk interval kepercayaannya:

$$\bar{x} - z_{\frac{1}{2}\gamma}.\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\frac{1}{2}\gamma}.\frac{\sigma}{\sqrt{n}}$$

b. Simpangan baku σ tidak diketahui dan populasi berdistribusi normal

$$P\left(\bar{x} - t_p.\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_p.\frac{s}{\sqrt{n}}\right) = \gamma$$

dengan γ = koefisien kepercayaan dan t_p = nilai t didapat dari daftar distribusi student dengan $p = \frac{1}{2}(1 + \gamma)$ dan $dk = (n - 1)$

Untuk interval kepercayaannya:

$$\bar{x} - t_p.\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_p.\frac{s}{\sqrt{n}}$$

c. Simpangan baku σ tidak diketahui dan populasi tidak berdistribusi normal

Jika n cukup besar maka dalil limit pusat berlaku maka dapat digunakan cara a. dengan menggunakan kekeliruan yang sangat kecil. Jika populasi sangat menyimpang dari normal dan ukuran sampel kecil sekali maka teorinya harus dipecahkan dengan menggunakan bentuk distribusi asli dari populasi bersangkutan.

Contoh:

Sebuah sampel acak terdiri dari 100 mahasiswa tealh diambil daris sebuah Universitas lain nilai-nilai IQ-nya dicatat. Didapat $\bar{x} = 112$ dan $s = 10$.

a) Kita dapat mengatakan: IQ rata-rata untuk mahasiswa Universitas itu = 112

Dalam hal ini digunakan titik taksiran.

b) Jika dikehendaki interval taksiran IQ rata-rata dengan koefisien kepercayaan 0,95 maka $p = \frac{1}{2}(1 + 0{,}95) = 0{,}975$ dan $dk = 100 - 1 = 99$ dengan menggunakan interpolasi dari Daftar G dalam lampiran didapat $t_p = 1{,}987$.
Maka interval kepercayaan

$$112 - (1{,}987)\frac{10}{\sqrt{100}} < \mu < 112 + (1{,}987)\frac{10}{\sqrt{100}}$$

Atau: $110 < \mu < 114$
Jadi didapat 95% interval kepercayaan untuk IQ rata-rata mahasiswa ialah $110 < \mu < 114$

2. Interval Kepercayaan bagi selisih rata-rata

Misalkan terdapat dua populasi, kedua-duanya berdistribusi normal. Rata-rata dan simpangan bakunya masing-masing μ_1 dan σ_1 untuk populasi kesatu, μ_2 dan σ_2 untuk populasi kedua. Dari masing-masing populasi secara independent diambil sebuah sampel acak dengan ukuran n_1 dan n_2. Rata-rata dan simpangan baku dari sampel-sampel itu berturut-turut $\overline{x_1}, s_1$ dan $\overline{x_2}, s_2$. Akan ditaksir selisih rata-rata $(\mu_1 - \mu_2)$.

Jelas bahwa titik taksiran untuk $(\mu_1 - \mu_2)$ adalah $(\overline{x_1} - \overline{x_2})$. Untuk memperoleh taksiran yang lebih tinggi derajat kepercayaan, digunakan interval taksiran atau selang taksiran disertai nilai koefisien kepercayaan yang dikehendaki. Dibedakan menjadi tiga hal

a. $\sigma_1 = \sigma_2$

Jika kedua populasi normal itu mempunyai $\sigma_1 = \sigma_2 = \sigma$ dan **besarnya diketahui**, maka dengan interval kepercayaan $\gamma\%$ untuk $(\mu_1 - \mu_2)$ ditentukan oleh rumus:

$$(\overline{x_1} - \overline{x_2}) - z_{\frac{1}{2}\gamma}.\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\overline{x_1} - \overline{x_2}) + z_{\frac{1}{2}\gamma}.\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Dengan $z_{\frac{1}{2}\gamma}$ didapat dari distribusi normal baku dengan peluang $\frac{1}{2}\gamma$.

Dalam hal $\sigma_1 = \sigma_2 = \sigma$ tetapi **tidak diketahui besarnya**, pertama-tama dari sampel-sampel perlu ditentukan varians gabungannya, dinyatakan dengan s^2, besarnya diberikan oleh rumus:

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Interval kepercayaannya ditentukan dengan menggunakan distribusi student. Formula dengan interval kepercayaan $\gamma\%$ untuk $(\mu_1 - \mu_2)$ adalah

$$(\overline{x_1} - \overline{x_2}) - t_p.s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\overline{x_1} - \overline{x_2}) + t_p.s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

dengan t_p didapat dari daftar distribusi student dengan $p = \frac{1}{2}(1 + \gamma)$ dan $dk = (n_1 + n_2 - 2)$

b. $\sigma_1 \neq \sigma_2$

Dengan memisalkan $s_1 = \sigma_1$ dan $s_2 = \sigma_2$, untuk sampel-sampel acak berukuran cukup besar, dapt dilakukan pendekatan kepada distribusi normal. Formula interval kepercayaannya ditentukan oleh:

$$(\overline{x_1} - \overline{x_2}) - z_{\frac{1}{2}\gamma}.\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\overline{x_1} - \overline{x_2}) + z_{\frac{1}{2}\gamma}.\sigma\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Contoh:
Ada dua cara pengukuran untuk mengukur kelembaman suatu zat. Cara I dilakukan 50 kali yang menghasilkan $\overline{x_1} = 60{,}2$ dan $s_1^2 = 24{,}7$. Cara II dilakukan 60 kali dengan $\overline{x_2} = 70{,}4$ dan $s_2^2 = 37{,}2$.
Supaya ditentukan interval kepercayaan 95% mengenai perbedaan rata-rata pengukuran dari kedua cara itu
Jawab:
Jika dimisalkan hasil kedua cara pengukuran berdistribusi normal, maka didapat varians gabungan:

$$s^2 = \frac{(50-1)(24{,}7) + (60-1)(37{,}2)}{50+60-2} = 31{,}53$$

Selanjutnya dihitung dulu:

$$s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = \sqrt{\frac{31{,}53}{50} + \frac{31{,}53}{60}} = 1{,}08$$

Dengan $p = 0{,}975$ dan $dk = 108$, dari daftar distribusi t didapat $t = 1{,}984$. Maka interval kepercayaan:

$$(70{,}4 - 60{,}2) - (1{,}984)(1{,}08) < \mu_1 - \mu_2 < (70{,}4 - 60{,}2) + (1{,}984)(1{,}08)$$

Atau

$$8{,}06 < \mu_1 - \mu_2 < 12{,}34$$

c. Observasi berpasangan
Misalkan populasi kesatu mempunyai variabel acak X dan populasi kedua mempunyai variabel acak Y. Rata-ratanya masing-masing μ_x dan μ_y. Diambil dua sampel acak masing-masing sebuah dari tiap populasi, yang berukuran sama, jadi $n_1 = n_2 = n$. Didapat data sampel: $(x_1, x_2, \dots, x_n)$ dan $(y_1, y_2, \dots, y_n)$. Kedua data hasil observasi ini dimisalkan berpasangan sebagai berikut:
x_1 berpasangan dengan y_1
x_2 berpasangan dengan y_2
$\vdots$
x_n berpasangan dengan y_n
Dalam hal pasangan data seperti ini, maka menaksir selisih atau beda rata-rata $\mu_B = \mu_x - \mu_y$, dapat pula dibentuk selisih atau beda tiap pasangan data. Jadi dicari $B_1 = x_1 - y_1$, $B_2 = x_2 - y_2$, …, $B_n = x_n - y_n$.
Dari sampel berukuran n yang datanya terdiri dari $B_1, B_2, \dots, B_n$ supaya dihitung rata-rata $\bar{B}$ dan simpangan baku s_B, dengan menggunakan:

$$\bar{B} = \frac{\sum B_i}{n} \text{ dan } s_B^2 = \frac{n\sum B_i^2 - (\sum B_i)^2}{n(n-1)}$$

Maka interval kepercayaan untuk μ_B dengan koefisien kepercayaan $\gamma\%$ yaitu:

$$\bar{B} - t_p.\frac{s_B}{\sqrt{n}} < \mu_B < \bar{B} + t_p.\frac{s_B}{\sqrt{n}}$$

Dengan t_p didapat dari daftar distribusi student untuk $p = \frac{1}{2}(1+\gamma)$ dan $dk = (n-1)$

3. Interval Kepercayaan bagi proporsi
Misalkan populasi berdistribusi binom berukuran N, terdapat proporsi π untuk suatu kejadian A dalam populasi tersebut. Diambil sampel acak berukuran n dari populasi itu dengan proporsi $\left(\frac{x}{n}\right)$ untuk kejadian A dalam sampel tersebut. Jadi taksiran titik untuk π adalah $\left(\frac{x}{n}\right)$. Maka interval kepercayaan untuk taksiran π dengan koefisien kepercayaan $\gamma\%$ yaitu:

$$\sum_{y=x}^{n} \binom{n}{y} \pi^y (1-\pi)^{n-y} = \frac{1}{2}(1-\gamma) \qquad \text{.....(A)}$$

$$\sum_{y=0}^{x} \binom{n}{y} \pi^y (1-\pi)^{n-y} = \frac{1}{2}(1-\gamma) \qquad \text{.....(B)}$$

Formula (A) merupakan batas bawah interval kepercayaan dan formula (B) merupakan batas atas interval kepercayaan.

Rumus diatas tidak praktis, sehingga sering kali digunakan pendekatan distribusi normal kepada binom untuk ukuran sampel n cukup besar. Maka interval kepercayaan π, dengan koefisien kepercayaan $\gamma\%$ adalah:

$$p - z_{\frac{1}{2}\gamma}\sqrt{\frac{pq}{n}} < \pi < p + z_{\frac{1}{2}\gamma}\sqrt{\frac{pq}{n}}$$

Dengan $p = \frac{x}{n}$ dan $q = 1 - p$

Contoh

Misalkan kita ingin menaksir ada berapa persen anggota masyarakat berumur 15 tahun ke atas yang termasuk ke dalam golongan A. Untuk ini sampel acak berukuran acak $n = 1200$ diambil yang menghasilkan 504 tergolong kategori A.

Jawab:

Persentase golongan A dalam sampel $= \frac{504}{1200} \times 100\% = 42\%$

Jika ditaksir ada 42% anggota masyarakat berumur 15 tahun ke atas yang termasuk golongan A, maka dalam hal ini telah digunakan titik taksiran. Untuk menentukan 95% interval kepercayaan parameter π, untuk n yang cukup besar, dengan $p = 0{,}42$; $q = 0{,}58$; $z_{0{,}475} = 1{,}96$, maka:

$$0{,}42 - (1{,}96)\sqrt{\frac{0{,}42 \times 0{,}58}{1200}} < \pi < 0{,}42 + (1{,}96)\sqrt{\frac{0{,}42 \times 0{,}58}{1200}}$$

Atau: $0{,}39 < \pi < 0{,}45$

4. Interval Kepercayaan bagi selisih proporsi

Misal terdapat dua populasi berdistribusi binom dengan parameter untuk peristiwa yang sama masing-masing π_1 dan π_2. Dari populasi ini secara independent masing-masing diambil sebuah sampel acak berukuran n_1 dari populasi kesatu dan n_2 dari populasi kedua. Proporsi untuk peristiwa yang diperhatikan dari sampel-sampel itu adalah $p_1 = \frac{x_1}{n_1}$ dan $p_2 = \frac{x_2}{n_2}$ dengan x_1 dan x_2 berturut-turut menyatakan banyaknya peristiwa yang diperhatikan yang terdapat pada sampel kesatu dan sampel kedua. Penentuan interval kepercayaan untuk $(\pi_1 - \pi_2)$ akan digunakan pendekatan oleh distribusi normal dengan koefisien kepercayaan $\gamma\%$, yaitu:

$$(p_1 - p_2) - z_{\frac{1}{2}\gamma}\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} < \pi_1 - \pi_2 < (p_1 - p_2) + z_{\frac{1}{2}\gamma}\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

Dengan $q_i = 1 - p_i$

Contoh

Misal sampel acak satu terdiri 500 wanita dan sampel acak kedua 700 laki-laki yang mengunjungi sebuah pameran telah diambil. Ternyata bahwa 325 wanita dan 400 laki-laki menyenangi pameran itu.

Tentukan interval kepercayaan 95% untuk perbedaan persentase laki-laki dan wanita yang mengunjungi pameran dan menyenanginya.

Jawab:

Peresntase wanita yang menyukai pameran $p_1 = \frac{325}{500} \times 100\% = 65\%$ dan untuk laki-laki $p_2 = \frac{400}{700} \times 100\% = 57\%$

Jadi $q_1 = 35\%$ dan $q_2 = 43\%$

Dengan $n_1 = 500$ dan $n_2 = 700$, didapat

$$\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} = \sqrt{\frac{0.65 \times 0.35}{500} + \frac{0.57 \times 0.43}{700}} = 0.0284$$

Dengan $z = 1{,}96$ diperoleh:

$$(0.65 - 0.57) - (1.96)(0.0284) < \pi_1 - \pi_2 < (0.65 - 0.57) + (1.96)(0.0284)$$

Atau: $0.024 < \pi_1 - \pi_2 < 0.136$

INTERPOLASI

Jika diketahui $t_{0,975,60} = 2$ dan $t_{0,975,120} = 1{,}98$, tentuka $t_{0,975,99}$?

Jawab:

Gunakan persamaan garis:

$$\frac{y - y_1}{y_2 - y_1} = \frac{x - x_1}{x_2 - x_1}$$

Dengan $(x_1, y_1) = (60{,}2)$ dan $(x_2, y_2) = (120{,}1.98)$ maka diperoleh

$$\frac{y - 2}{1.98 - 2} = \frac{x - 60}{120 - 60}$$

$$\frac{y - 2}{-0.02} = \frac{x - 60}{60}$$

$$y - 2 = -\frac{0{,}02}{60}x + 0{,}02$$

$$y = -\frac{0{,}02}{60}x + 2{,}02$$

Substitusi x = 99, maka diperoleh

$$y = -\frac{0{,}02}{60}(99) + 2{,}02 = 1.987$$

Maka $t_{0,975,99} = 1.987$