Kumpulan Soal & Pembahasan: Estimasi Parameter & Interval Kepercayaan

10 Soal Terapan Rekayasa & Sains Data: Selang Z, Distribusi t-Student, Proporsi, dan Ukuran Sampel Minimum (Slide Soal & Pembahasan Terpisah)

Statistika dan Probabilitas · Pertemuan 09

Daftar 10 Soal Terapan Pertemuan 09

Bagian A: Interval Z (σ Diketahui)

- **Soal 1:** Uji Kuat Tarik Kabel Fiber Optik (95% CI)
- **Soal 2:** Komparasi Tingkat Keyakinan 90%, 95%, 99%

Bagian B: Distribusi t (σ Tak Diketahui)

- **Soal 3:** Latensi Query Database Cluster ($n = 16$)
- **Soal 4:** Daya Tahan Baterai Sensor IoT ($n = 25$)

Bagian C: Selang Proporsi ($\hat{p}$)

- **Soal 5:** Rasio Klik Iklan Digital (A/B Testing)
- **Soal 6:** Audit Cacat Mutu Modul Kamera HP
- **Soal 7:** Uji Validitas Syarat Normalitas Proporsi

Bagian D: Ukuran Sampel Minimum (n)

- **Soal 8:** Alokasi Sampel Uji Beban Server Cloud
- **Soal 9:** Sampel Survei Quick Count Pilkada
- **Soal 10:** Perencanaan Sampel Konservatif vs Pilot

Bagian A — Selang Kepercayaan Z (σ Diketahui)

Penerapan Kurva Normal Baku saat Variabilitas Populasi Stabil

Soal 1: Uji Kuat Tarik Kabel Fiber Optik

KASUS

Sebuah pabrik kabel fiber optik memiliki standar kendali mutu dengan simpangan baku kekuatan tarik kabel yang diketahui stabil sebesar $\sigma = 4{,}0$ MPa.

Tim inspeksi mengambil sampel acak $n = 64$ segmen kabel dari batch produksi terbaru dan memperoleh rata-rata kekuatan tarik sampel $\overline{x} = 125{,}5$ MPa.

Pertanyaan

1. Hitung **Standard Error** (SE) dan **Margin of Error** (E) untuk tingkat kepercayaan 95% ($z_{0{,}025} = 1{,}96$)!
2. Konstruksikan selang kepercayaan 95% untuk rata-rata kekuatan tarik seluruh kabel dalam batch produksi tersebut!

Pembahasan Soal 1: Perhitungan Selang Z 95%

Langkah 1: Identifikasi Parameter & Standard Error

Diketahui: $\overline{x} = 125{,}5$, $\sigma = 4{,}0$, $n = 64$, $z_{0{,}025} = 1{,}96$.

$$\text{SE} = \frac{\sigma}{\sqrt{n}} = \frac{4{,}0}{\sqrt{64}} = \frac{4{,}0}{8} = 0{,}50\,\text{MPa}$$

Langkah 2: Margin of Error & Interval Kepercayaan

$$E = z_{\frac{\alpha}{2}} \times \text{SE} = 1{,}96 \times 0{,}50 = 0{,}98\,\text{MPa}$$

$$\text{CI}_{95\%} = [\overline{x} - E, \overline{x} + E] = [125{,}5 - 0{,}98, 125{,}5 + 0{,}98] = [124{,}52, 126{,}48]\,\text{MPa}$$

Interpretasi Ilmiah: Kita 95% yakin bahwa rata-rata kekuatan tarik populasi kabel fiber optik batch tersebut berada di antara 124,52 MPa hingga 126,48 MPa.

Soal 2: Komparasi Efek Tingkat Keyakinan (90%, 95%, 99%)

KASUS

Menggunakan data dari Soal 1 ($\overline{x} = 125{,}5$ MPa, $\sigma = 4{,}0$ MPa, $n = 64$, SE $= 0{,}50$ MPa):

Pertanyaan

1. Hitung selang kepercayaan untuk tingkat keyakinan 90% ($z_{0{,}05} = 1{,}645$)!
2. Hitung selang kepercayaan untuk tingkat keyakinan 99% ($z_{0{,}005} = 2{,}576$)!
3. Bandingkan lebar ketiga interval (90%, 95%, 99%). Apa kesimpulan penting mengenai trade-off antara tingkat kepastian dan presisi interval?

Pembahasan Soal 2: Komparasi Lebar Selang Kepercayaan

Perhitungan Interval 90% dan 99%

- **Tingkat 90% ($z = 1{,}645$):**

$E_{90\%} = 1{,}645 \times 0{,}50 = 0{,}8225$ MPa

$\text{CI}_{90\%} = [125{,}5 - 0{,}8225, 125{,}5 + 0{,}8225] = [124{,}68, 126{,}32]$ MPa (Lebar $= 1{,}645$)

- **Tingkat 99% ($z = 2{,}576$):**

$E_{99\%} = 2{,}576 \times 0{,}50 = 1{,}288$ MPa

$\text{CI}_{99\%} = [125{,}5 - 1{,}288, 125{,}5 + 1{,}288] = [124{,}21, 126{,}79]$ MPa (Lebar $= 2{,}576$)

Kesimpulan Trade-off: Semakin tinggi tingkat keyakinan yang diinginkan ($90\% \rightarrow 95\% \rightarrow 99\%$), selang interval menjadi semakin LEBAR ($1{,}645 \rightarrow 1{,}960 \rightarrow 2{,}576$ MPa). Kita memperoleh kepastian lebih besar, namun presisi estimasi menurun!

Bagian B — Distribusi t-Student (σ Tak Diketahui)

Estimasi Realistis dengan Sampel Kecil / Menengah dan Ragam Sampel s

Soal 3: Waktu Eksekusi Kueri Database Cluster

KASUS

Seorang database administrator mengukur waktu eksekusi kueri agregasi SQL kompleks pada cluster database. Simpangan baku populasi (σ) tidak diketahui.

Administrator mengambil sampel acak $n = 16$ kali eksekusi dan mencatat rata-rata sampel $\overline{x} = 320$ ms dengan simpangan baku sampel $s = 40$ ms. Waktu kueri diasumsikan berdistribusi normal.

Pertanyaan

Tentukan derajat bebas (df), nilai kritis t, dan susun selang kepercayaan 95% untuk rata-rata waktu kueri populasi! (Diketahui tabel: $t_{0,025,15} = 2{,}131$).

Pembahasan Soal 3: Selang Kepercayaan t-Student 95%

Langkah 1: Identifikasi Derajat Bebas & Nilai Kritis

Derajat bebas: $df = n - 1 = 16 - 1 = 15$.

Karena σ tidak diketahui dan $n = 16 < 30$, wajib memakai distribusi t: $t_{0{,}025,15} = 2{,}131$.

Langkah 2: Margin of Error & Interval

$$E = t_{\frac{\alpha}{2},\,15} \times \frac{s}{\sqrt{n}} = 2{,}131 \times \frac{40}{\sqrt{16}} = 2{,}131 \times 10 = 21{,}31\ \text{ms}$$

$$\text{CI}_{95\%} = [320 - 21{,}31,\ 320 + 21{,}31] = [298{,}69,\ 341{,}31]\ \text{ms}$$

Catatan Kritis: Jangan menggunakan $z = 1{,}96$. Menggunakan z pada sampel kecil dengan s mengabaikan variabilitas penaksir simpangan baku, yang menghasilkan interval yang terlalu sempit secara keliru.

Soal 4: Daya Tahan Baterai Sensor IoT Pertanian

KASUS

Sebuah sensor kelembaban tanah nirkabel (IoT) diuji ketahanan baterainya. Dari sampel acak $n = 25$ sensor, rata-rata masa pakai baterai tercatat $\overline{x} = 180$ hari dengan simpangan baku sampel $s = 20$ hari.

Pertanyaan

1. Tentukan derajat bebas dan nilai kritis tabel t untuk tingkat kepercayaan 99%! (Diketahui: $t_{0{,}005,24} = 2{,}797$).
2. Hitung selang kepercayaan 99% untuk rata-rata daya tahan baterai sensor tersebut!
3. Apakah produsen aman mengklaim bahwa baterai sensor mampu bertahan minimal 165 hari?

Pembahasan Soal 4: Selang t 99% dan Evaluasi Garansi

Perhitungan Margin of Error & Interval 99%

Derajat bebas: $df = 25 - 1 = 24$. Nilai kritis: $t_{0{,}005,24} = 2{,}797$.

$$E = 2{,}797 \times \frac{20}{\sqrt{25}} = 2{,}797 \times 4 = 11{,}188 \,\text{hari}$$

$$\text{CI}_{99\%} = [180 - 11{,}19,\, 180 + 11{,}19] = [168{,}81,\, 191{,}19] \,\text{hari}$$

Evaluasi Klaim Produsen (Garansi 165 Hari)

- Batas bawah selang kepercayaan 99% adalah 168,81 hari.
- Karena angka 165 hari berada di bawah batas terendah selang ($165 < 168{,}81$), produsen **sangat aman dan terbukti secara ilmiah** untuk menggaransikan daya tahan minimal 165 hari.

Bagian C — Selang Kepercayaan Proporsi ($\hat{p}$)

Inferensi Data Kategorik Biner (Sukses / Gagal, Konversi, Cacat)

Soal 5: Rasio Klik Iklan Digital (A/B Testing CTR)

KASUS

Sebuah platform e-commerce meluncurkan desain baru tombol **Checkout**. Dari sampel acak $n = 400$ pengunjung yang melihat desain baru, sebanyak $X = 36$ orang mengklik tombol tersebut.

Pertanyaan

1. Hitung proporsi sampel ($\hat{p}$) dan periksa kelayakan syarat normalitas ($n\hat{p} \geq 10$ dan $n(1-\hat{p}) \geq 10$)!
2. Hitung selang kepercayaan 95% ($z = 1{,}96$) untuk proporsi klik populasi sesungguhnya (p)!

Pembahasan Soal 5: Estimasi Selang Proporsi

Langkah 1: Proporsi Sampel & Uji Syarat Normal

$\hat{p} = \frac{36}{400} = 0{,}09$ (9%).

Uji syarat: $n\hat{p} = 400 \times 0{,}09 = 36 \geq 10$ dan $n(1-\hat{p}) = 400 \times 0{,}91 = 364 \geq 10$. Syarat terpenuhi!

Langkah 2: Margin of Error & Interval Proporsi

$$\text{SE} = \sqrt{\frac{0{,}09 \times 0{,}91}{400}} = \sqrt{\frac{0{,}0819}{400}} = \sqrt{0{,}00020475} \approx 0{,}01431$$

$$E = 1{,}96 \times 0{,}01431 = 0{,}0280\,(2{,}80\%)$$

$$\text{CI}_{95\%} = [0{,}09 - 0{,}0280,\, 0{,}09 + 0{,}0280] = [0{,}0620,\, 0{,}1180] \quad \rightarrow \quad [6{,}20\%,\, 11{,}80\%]$$

Soal 6: Audit Cacat Mutu Modul Kamera Smartphone

KASUS

Pabrik perakitan smartphone mengaudit cacat sensor kamera. Dari $n = 600$ unit modul kamera yang diperiksa secara acak, ditemukan $X = 18$ unit mengalami cacat fokus.

Pertanyaan

1. Hitung estimasi titik proporsi cacat ($\hat{p}$)!
2. Konstruksikan selang kepercayaan 90% ($z_{0,05} = 1{,}645$) untuk proporsi cacat seluruh produksi kamera!
3. Jika standar toleransi cacat industri adalah maksimal 4%, apakah pabrik ini berada dalam batas aman?

Pembahasan Soal 6: Audit Cacat Kamera HP

Perhitungan Interval Proporsi 90%

$\hat{p} = \frac{18}{600} = 0{,}03$ (3%). Syarat: $600 \times 0{,}03 = 18 \geq 10$ dan $600 \times 0{,}97 = 582 \geq 10$ (Valid).

$$E = 1{,}645 \times \sqrt{\frac{0{,}03 \times 0{,}97}{600}} = 1{,}645 \times \sqrt{0{,}0000485} = 1{,}645 \times 0{,}006964 \approx 0{,}0115 \, (1{,}15\%)$$

$$\text{CI}_{90\%} = [0{,}03 - 0{,}0115, 0{,}03 + 0{,}0115] = [0{,}0185, 0{,}0415] \quad \rightarrow \quad [1,85\%, 4,15\%]$$

Evaluasi Toleransi Cacat 4%

Peringatan Mutu: Batas atas selang mencapai $4,15\%$, yang melampaui toleransi 4%. Meskipun estimasi titiknya 3%, pabrik belum bisa 90% yakin bahwa produksinya aman di bawah 4%. Disarankan investigasi kalibrasi mesin!

Soal 7: Uji Batasan Syarat Normalitas Proporsi

KASUS

Seorang peneliti keamanan siber menguji keberhasilan eksploitasi celah keamanan baru. Peneliti melakukan $n = 50$ kali percobaan injeksi dan berhasil menembus sistem sebanyak $X = 3$ kali ($\hat{p} = \frac{3}{50} = 0{,}06$).

Pertanyaan

1. Apakah rumus selang kepercayaan proporsi berbasis distribusi normal (z) sah digunakan pada data ini? Buktikan dengan perhitungan syarat kelayakan!
2. Apa metode yang tepat digunakan jika syarat normalitas tidak terpenuhi?

Pembahasan Soal 7: Mengapa CLT Gugur pada Sampel Jarang?

1. Pengujian Syarat Normalitas

$$n\hat{p} = 50 \times 0{,}06 = 3 < 10 \quad \textbf{(GUGUR / TIDAK TERPENUHI!)}$$

Meskipun $n(1-\hat{p}) = 50 \times 0{,}94 = 47 \geq 10$, syarat $n\hat{p} \geq 10$ tidak terpenuhi karena jumlah kejadian sukses terlalu sedikit. Distribusi sampling sangat miring (**skewed**) dan pendekatan kurva normal Gauss menjadi tidak valid.

2. Alternatif Metodologis yang Sah

- **Metode Eksak Clopper-Pearson:** Menghitung interval menggunakan fungsi kumulatif distribusi Binomial eksak.
- **Koreksi Wilson Score Interval:** Memberikan performa yang jauh lebih baik untuk proporsi ekstrem dekat 0 atau 1 pada sampel kecil.

Bagian D — Penentuan Ukuran Sampel Minimum (n)

Perencanaan Efisiensi Anggaran dan Target Presisi Rekayasa

Soal 8: Alokasi Sampel Uji Beban Server Cloud

KASUS

Arsitek cloud ingin mengestimasi rata-rata latensi API gateway sistem perbankan. Berdasarkan data historis pengujian bulan lalu, simpangan baku latensi stabil pada $\sigma = 30$ ms.

Manajemen menginginkan estimasi memiliki **Margin of Error** tidak lebih dari $E = 4$ ms pada tingkat keyakinan 95% ($z = 1{,}96$).

Pertanyaan

Berapa ukuran sampel request minimal (n) yang harus dikumpulkan oleh arsitek cloud tersebut?

Pembahasan Soal 8: Perhitungan Ukuran Sampel Rata-Rata

Formula Ukuran Sampel Rata-Rata: $n = (\frac{z_{\frac{\alpha}{2}}\,\sigma}{E})^2$

Diketahui: $\sigma = 30$, $E = 4$, $z_{0{,}025} = 1{,}96$.

$$n = (\frac{1{,}96 \times 30}{4})^2 = (\frac{58{,}8}{4})^2 = (14{,}7)^2 = 216{,}09$$

Aturan Pembulatan ke Atas (Ceiling Rule)

Karena ukuran sampel harus berupa bilangan bulat, dan pembulatan ke bawah (216) akan membuat $E > 4$ ms, maka nilai n wajib dibulatkan ke atas:

$$n = 217 \,\text{request}$$

Soal 9: Perencanaan Sampel Quick Count Pemilu

KASUS

Lembaga survei independen merencanakan survei **Quick Count** pada Pilkada Gubernur. Lembaga menargetkan **Margin of Error** maksimal $E = 0{,}015$ (1,5%) pada tingkat kepercayaan 95% ($z = 1{,}96$).

Karena pemilihan diprediksi sangat ketat dan belum ada data awal yang pasti, lembaga mengadopsi asumsi paling konservatif (**worst-case scenario**).

Pertanyaan

Berapa banyak pemilih / TPS yang harus disurvei untuk memenuhi target presisi ketat tersebut?

Pembahasan Soal 9: Ukuran Sampel Proporsi Konservatif

Asumsi Konservatif ($p^* = 0{,}50$)

Nilai variabilitas maksimum terjadi saat $p^*(1-p^*) = 0{,}50 \times 0{,}50 = 0{,}25$.

$$n = \left(\frac{z_{\frac{\alpha}{2}}}{E}\right)^2 p^*(1-p^*) = \left(\frac{1{,}96}{0{,}015}\right)^2 \times 0{,}25$$

$$n = (130{,}6667)^2 \times 0{,}25 = 17.073{,}78 \times 0{,}25 = 4.268{,}44$$

Kesimpulan Pembulatan

$$n = 4.269 \,\text{responden}$$

Lembaga survei wajib mengumpulkan minimal 4.269 sampel untuk menjamin MoE $\leq 1{,}5\%$.

Soal 10: Efisiensi Biaya Sampel: Pilot Study vs Asumsi Konservatif

KASUS

Startup fintech ingin mengetahui proporsi pengguna yang berminat pada fitur **PayLater**. Target $E = 0{,}03$ (3%) pada tingkat keyakinan 95% ($z = 1{,}96$).

Pertanyaan

1. Hitung ukuran sampel jika menggunakan pendekatan konservatif ($p^* = 0{,}50$)!
2. Sebelum survei besar, tim melakukan studi perintis (**pilot study**) kecil dan mendapati dugaan minat awal hanya $p^* = 0{,}15$. Hitung ukuran sampel baru menggunakan informasi ini!
3. Berapa persen biaya/responden yang berhasil dihemat berkat studi perintis tersebut?

Pembahasan Soal 10: Nilai Ekonomis Pilot Study

1. Konservatif vs 2. Pilot Study

- **Konservatif ($p^* = 0{,}50$):**

$n_1 = (\frac{1{,}96}{0{,}03})^2(0{,}25) = (65{,}33)^2 \times 0{,}25 = 4.268{,}44 \times 0{,}25 = 1.067{,}11 \rightarrow 1.068$ orang.

- **Pilot Study ($p^* = 0{,}15$):**

$p^*(1-p^*) = 0{,}15 \times 0{,}85 = 0{,}1275$

$n_2 = 4.268{,}44 \times 0{,}1275 = 544{,}22 \rightarrow 545$ orang.

3. Analisis Penghematan Biaya Riset

$$\text{Penghematan} = \frac{1.068-545}{1.068} \times 100\% = \frac{523}{1.068} \times 100\% \approx 48{,}97\%$$

Pelajaran Data Science: Melakukan studi perintis kecil ($n = 50$) terlebih dahulu berhasil memangkas hampir 50% kebutuhan sampel survei utama!