BAB VIII
MASALAH ESTIMASI SATU DAN DUA SAMPEL

8.1 Statistik inferensial

Statistik inferensial suatu metode mengambil kesimpulan dari suatu populasi. Ada dua pendekatan yang digunakan dalam statistik inferensial. Pertama, **metode klasik**, yaitu inferensi didasarkan hanya pada informasi yang didapatkan dari sampel acak yang dipilih dari populasi. Kedua, **metode Bayesian**, yaitu dengan menggunakan pengetahuan sebelumnya tentang distribusi probabilitas dari parameter yang tidak dketahui terkait dengan data sampel. Statistik inferensial digunakan untuk dua hal utama, yaitu **Estimasi** dan **uji hipotesis**.

8.2 Estimasi dengan metode klasik

Point estimate dari suatu parameter populasi θ adalah suatu nilai tunggal $\hat{\theta}$ dari suatu statistik $\hat{\Theta}$. Misal: nilai $\bar{x}$ dari statistik $\bar{X}$ yang dihitung dari n sampel adalah suatu titik perkiraan dari parameter populasi μ.

<u>Unbiased estimator</u>

Definisi: Statistik $\hat{\Theta}$ dikatakan unbiased estimator dari parameter θ bila

$$\mu_{\hat{\Theta}} = E(\hat{\Theta}) = \theta$$

<u>Varians dari point estimator</u>

Bila $\hat{\Theta}_1$ dan $\hat{\Theta}_2$ adalah dua unbiased estimator dari parameter populasi yang sama θ, maka yang dipilih adalah estimator yang menghasilkan varians lebih kecil.

Estimasi Interval

Estimasi interval parameter populasi θ adalah interval dalam bentuk $\hat{\theta}_L < \theta < \hat{\theta}_U$,

dimana $\hat{\theta}_L$ dan $\hat{\theta}_U$ tergantung pada nilai statistik $\hat{\Theta}$ untuk sampel tertentu dan juga pada sampling distribution $\hat{\Theta}$.

Interpretasi Estimasi Interval

Karena sampel yang berbeda akan menghasilkan nilai $\hat{\Theta}$ yang berbeda dan juga nilai $\hat{\theta}_L$ dan $\hat{\theta}_U$ yang berbeda. Dari distribusi sampling $\hat{\Theta}$ kita harus dapat menentukan $\hat{\Theta}_L$ dan $\hat{\Theta}_U$ sedemikian rupa sehingga:

$$P(\hat{\Theta}_L < \theta < \hat{\Theta}_U) = 1 - \alpha \quad \text{untuk } 0 < \alpha < 1,$$

Interval $\hat{\theta}_L < \theta < \hat{\theta}_U$ yang dihitung dari sampel yang dipilih disebut (1- α)100% **interval keyakinan** (*confidence interval*). 1- α dinamakan **koefesien keyakinan** (*confidence coefficient*) atau **tingkat keyakinan** (*degree of confidence*), dan $\hat{\theta}_L$ dan $\hat{\theta}_U$ dsiebut **batas keyakinan** (*confidence limit*). Sehingga bila α = 0.05; berarti interval keyakinan adalah 95%. Semakin lebar interval keyakinan semakin yakin dalam penentuan parameter.

8.3 Sampel tunggal: menaksir rata-rata

Bila $\bar{x}$ adalah rata-rata dari *n* sampel acak dari populasi dengan varians σ^2, interval keyakinan (1- α)100% untuk μ adalah:

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$z_{\alpha/2}$ adalah nilai z yang terletak disisi kanan α/2.

Bila jumlah sampel kurang dari 30, hasil yang diperoleh tidak akurat. Sebaliknya bila sampel lebih besar dari 30 hasil yang didapat terjamin lebih akurat.

Contoh: rata-rata konsentrasi nikel dari 36 lokasi sampel sungai adalah 2.6 gram/mm. Tentukan interval keyakinan 95% dan 99% untuk rata-rata konsentrasi nikel. Asumsikan deviasi standar populasi adalah 0.3.

Solusi:

Titik perkiraan μ adalah $\bar{x}$ = 2.6. Nilai $z_{\alpha/2} = z_{0.025} = 1.96$ (lihat tabel distribusi normal). Maka pada interval keyakinan 95% adalah

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} =$$

$$2.6 - (1.96)\left(\frac{0.3}{\sqrt{36}}\right) < \mu < 2.6 + (1.96)\left(\frac{0.3}{\sqrt{36}}\right) = 2.50 < \mu < 2.70$$

pada interval keyakinan 99%; $z_{\alpha/2} = z_{0.005} = 2.575$

$$2.6 - (2.575)\left(\frac{0.3}{\sqrt{36}}\right) < \mu < 2.6 + (2.575)\left(\frac{0.3}{\sqrt{36}}\right) = 2.47 < \mu < 2.73$$

Dari hasil diatas menunjukkan dibutuhkan interval yang lebih lebar untuk meningkatkan tingkat keyakinan.

Teorema: bila $\bar{x}$ digunakan untuk menaksir μ, kita dapat yakin (1- α)100% bahwa kesalahan tidak melebihi $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.

Dari contoh, sebelumnya kita yakin 95% bahwa rata-rata sampel $\bar{x}$ =2.6 berbeda dari rata-rata sebenarnya μ sebesar tidak lebih dari 0.1.

Teorema: bila $\bar{x}$ digunakan untuk menaksir μ, kita dapat yakin (1- α)100% bahwa kesalahan tidak melebihi jumlah tertentu *e* bila jumlah sampel adalah:

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

Contoh: Berapa banyak sampel diperlukan pada contoh sebelumnya bila kita menghendaki tingkat keyakinan 95% dengan tingkat kesalahan kurang dari 0.05.

Solusi:

$$n = \left(\frac{z_{\alpha/2}\sigma}{e}\right)^2 = \left(\frac{(1.96)(0.3)}{0.05}\right)^2 = 138.3$$

KASUS: σ tidak diketahui

Sering dalam beberapa kasus, kita mengestimasi rata-rata populasi dengan varians yang tidak diketahui. Bila kita ambil sampel acak dengan distribusi normal, maka variabel acak,

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

adalah *Student's t-distribution* dengan derajat kebebasan n -1. *S* = standar deviasi sampel. *T* dapat digunakan untuk membuat interval keyakinan. Prosedur sama dengan sebelumnya, hanya σ diganti dengan *S* dan ditribusi normal standar diganti dengan distribusi *t*. Sehingga kita dapat menyatakan bahwa:

$$P(-t_{\alpha/2} < T < t_{\alpha/2}) = 1 - \alpha$$

$$P\left(-t_{\alpha/2} < \frac{\bar{X} - \mu}{S/\sqrt{n}} < t_{\alpha/2}\right) = 1 - \alpha$$

$$P\left(\bar{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

<u>Interval keyakinan pada sampel besar</u>

Bila n ≥ 30, maka s dapat menggantikan σ dan interval keyakinannya adalah:

$$\bar{x} \pm z_{\alpha/2}\frac{s}{\sqrt{n}}$$

8.4 Kesalahan standar (*standard error*)

Standard deviasi dari $\overline{X}$ merupakan *standard error* dari $\overline{X}$, dinyatakan sebagai adalah $\sigma/\sqrt{n}$. Sehingga batas keyakinannya (confidence limit) adalah:

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \quad \text{dan ditulis sebagai} \quad \bar{x} \pm z_{\alpha/2}\text{s.e.}(\bar{x})$$

Untuk kasus σ tidak diketahui, maka:

$$\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}} \quad \text{dan ditulis sebagai} \quad \bar{x} \pm t_{\alpha/2}\text{s.e.}(\bar{x})$$

8.5 Prediksi interval

Dalam beberapa kasus seorang peneliti ingin memprediksi kemungkinan nilai pengamatan akan datang (*value of a future observation*) dengan menggunakan nilai pengamatan sebelumnya. Prediksi interval untuk suatu pengamatan akan datang dengan tingkat keyakinan (1- α)100% dari suatu pengengamatan sebelumnya yang terdistribusi normal dengan rata-rata μ yang tidak diketahui dan varians diketahui σ^2 adalah :

$$\bar{x} - z_{\alpha/2}\sigma\sqrt{1+1/n} < x_0 < \bar{x} + z_{\alpha/2}\sigma\sqrt{1+1/n}$$

Contoh:

Karena penurunan suku bunga perbankan, suatu Bank banyak menerima aplikasi penggadaian. Berdasarkan pengalaman 50 peminjaman penggadaian sebelumnya menunjukkan bahwa rata-ratanya = 128,300 $. Asumsikan deviasi standar = 15,000 \$. Tentukan 95% prediksi interval jumlah pinjaman penggadaian.

Solusi:

$$\bar{x} - z_{\alpha/2}\sigma\sqrt{1+1/n} < x_0 < \bar{x} + z_{\alpha/2}\sigma\sqrt{1+1/n} =$$

$$128300 - (1.96)(15000)\sqrt{1+1/50} < x < 128300 + (1.96)(15000)\sqrt{1+1/50} =$$

$$98607 < x < 157992$$

Untuk kasus σ tidak diketahui, formulasi dinyatakan sebagai berikut:

$$\bar{x} - t_{\alpha/2}s\sqrt{1+1/n} < x_0 < \bar{x} + t_{\alpha/2}s\sqrt{1+1/n}$$

8.6 Mengestimasi perbedaan dua rata-rata

Bila $\bar{x}_1$ dan $\bar{x}_2$ adalah rata-rata sampel acak bebas n_1 dan n_2 dari suatu populasi dengan varians σ_1^2 dan σ_2^2 . Interval dengan tingkat keyakinan (1- α)100% untuk μ_1 - μ_2 dinyatakan sebagai:

$$(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Contoh:

Suatu percobaan tentang kebutuhan bahan bakar dilakukan pada dua tipe mesin, A dan B. 50 percobaan dilakukan pada mesin tipe A, dan 75 untuk tipe B. Rata-rata mesin A dapat menempuh jarak 36 miles untuk setiap galon bahan bakar, sedangkan untuk mesin B adalah 42 miles per galon. Tentukan μ_1 - μ_2 dengan tingkat keyakinan 96% . Asumsikan deviasi standar A dan B adalah 6 dan 8.

Solusi:

pada interval keyakinan 96%; $z_{\alpha/2} = z_{0.02} = 2.05$ (lihat tabel).

$$(42-36) - 2.05\sqrt{\frac{64}{75} + \frac{36}{50}} < \mu_1 - \mu_2 < (42-36) + 2.05\sqrt{\frac{64}{75} + \frac{36}{50}} =$$

$$3.43 < \mu_1 - \mu_2 < 8.57$$