Statistika dan Probabilitas
Materi 12: **Teori Sampling**

Haryadi

Universitas Muhammadiyah Palangka Raya
May 24, 2026

1 Sampel Random

Di dalam aplikasi sering dilakukan pengambilan kesimpulan dari suatu kelompok individu atau *populasi*. Karena alasan tertentu, peneliti tidak mungkin untuk mengamati seluruh anggota populasi, namun hanya mengamati bagian dari populasi yang dinamakan *sampel*. **Inferensi statistik** adalah pengambilan kesimpulan berdasarkan sampel.

Diketahui X variabel random dengan distribusi $f(x)$. Jika dilakukan percobaan atau pengukuran sebanyak n kali, maka diperoleh variabel random $X_1, X_2, \cdots, X_n$ yang dinamakan *sampel random* dari distribusi $f(x)$. Jika nilai-nilai sampel random tersebut berturut-turut adalah $x_1, x_2, \cdots, x_n$, maka nilai-nilai ini dinamakan nilai eksperimen atau *data sampel*. Karena pengukuran dapat dilakukan pada kondisi yang sama maka dapat diasumsikan variabel random $X_1, X_2, \cdots, X_n$ independen dengan distribusi masing-masing $f(x_1), f(x_2), \cdots, f(x_n)$ dan fungsi peluang bersama $f(x_1, x_2, \cdots, x_n) = f(x_1)f(x_2)\cdots f(x_n)$.

Definisi 1. *Diketahui $X_1, X_2, \cdots, X_n$ variabel random independen yang masing memiliki distribusi peluang $f(x)$. Variabel random $X_1, X_2, \cdots, X_n$ dinamaan* **sampel random** *berkuran n dari populasi $f(x)$ dan fungsi distribusi bersamanya dituliskan*

$$f(x_1, x_2, \cdots, x_n) = f(x_1)f(x_2)\cdots f(x_n).$$

Suatu kuantitas yang dihitung dari data sampel dinamakan **statistik**, sedangkan suatu kuantitas yang dimiliki oleh suatu populasi dinamakan **parameter**. Bisa terjadi nilai parameter suatu populasi diketahui atau tidak diketahui.

Contoh 1. Misalkan variabel random X menyatakan usia layar LCD yang diproduksi suatu perusahaan yang diasumsikan berdistribusi normal dengan mean μ dan varian σ^2 yang tidak diketahui. Satu-satunya cara untuk memperoleh informasi tentang μ dan σ^2 adalah dengan melakukan eksperimen random. Misalkan dilakukan eksperimen dengan mengambil secara random sebanyak $n = 100$ layar LCD, dan usia layar LCD yang tercatat adalah $X_1, X_2, X_3, \cdots, X_{100}$. Dalam hal ini $X_1, X_2, X_3, \cdots, X_{100}$ merupakan sampel random yang berasal dari distribusi normal tersebut. Ke 100 layar LCD tersebut dapat digunakan untuk memperoleh informasi tentang μ dan σ^2. Ukuran yang diperoleh dari ke 100 layar LCD tersebut merupakan statistik, sedangkan μ dan σ^2 merupakan parameter.

Misalkan suatu populasi memiliki distribusi tertentu dengan **mean populasi** μ dan **varian populasi** σ^2. Suatu sampel random $X_1, X_2, \cdots, X_n$ diambil dari populasi tersebut. Ada dua statistik yang penting, yaitu mean sampel dan varian sampel.

Definisi 2. *Diketahui $X_1, X_2, \cdots, X_n$ adalah sampel random berukuran n.*

(a) ***Mean sampel*** $\overline{X}$ *didefinisikan*

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

(b) ***Varian sampel*** S^2 *didefinisikan*

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2,$$

dan $S = \sqrt{S^2}$ dinamakan deviasi standar sampel.

Contoh 2. Suatu eksperimen random telah dilakukan sebanyak 5 kali dan diperoleh $X_1 = 3.5$, $X_2 = 3.2$, $X_3 = 3.4$, $X_4 = 3.3$ dan $X_5 = 3.6$. Carilah mean sampel dan varian sampelnya.

Penyelesaian. Mean sampelnya adalah

$$\overline{X} = \sum_{i=1}^{5} \frac{X_i}{5} = \frac{3.5 + 3.2 + 3.4 + 3.3 + 3.6}{5} = 3.4,$$

dan varian sampelnya adalah

$$\begin{aligned} S^2 &= \sum_{i=1}^{5} \frac{(X_i - \overline{X})^2}{5-1} \\ &= \frac{(3.5-3.4)^2 + (3.2-3.4)^2 + (3.4-3.4)^2 + (3.3-3.4)^2 + (3.6-3.4)^2}{4} \\ &= 0.025. \end{aligned}$$

Dengan demikian deviasi standar sampel adalah $S = \sqrt{0.025} = 0.158$.

2 Distribusi Sampel

Statistik merupakan random variabel yang tergantung pada hasil observasi sampel. Sementara itu sampel diambil dari populasi yang memiliki distribusi tertentu. Oleh karena itu statistik juga memiliki distribusi peluang terntentu.

Definisi 3. **Distribusi sampel** *adalah distribusi peluang suatu statistik.*

Sebagai contoh, distribusi peluang $\overline{X}$ dinamakan distribusi mean sample, distribusi peluang σ^2 dinamakan distribusi varian sampel. Distribusi sampel tergantung pada distribusi populasi, ukuran sampel dan cara sampel diambil.

Teorema 1. *Jika $X_1, X_2, \cdots, X_n$ sampel random berukuran n dari suatu distribusi dengan mean μ dan varian σ^2, maka* *(a) nilai harapan $\overline{X}$ adalah $E[\overline{X}] = \mu$*

(b) varian $\overline{X}$ adalah $Var(\overline{X}) = \dfrac{\sigma^2}{n}$.

Bukti. (a) Karena untuk setiap X_i, $E[X_i] = \mu$, maka

$$\begin{aligned} E[\overline{X}] &= E\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right] \\ &= \frac{1}{n}(E[X_1] + E[X_2] + \cdots + E[X_n]) \\ &= \frac{1}{n}(\mu + \mu + \cdots + \mu) \\ &= \mu. \end{aligned}$$

(b) Dengan asumsi $X_1, X_2, \cdots, X_n$ independen,

$$\begin{aligned} Var(\overline{X}) &= Var\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) \\ &= \frac{1}{n^2}(Var(X_1) + Var(X_2) + \cdots + Var(X_n)) \\ &= \frac{n\sigma^2}{n^2} \\ &= \frac{\sigma^2}{n}. \end{aligned}$$

□

Berdasarkan teorema di atas, nilai harapan $\overline{X}$ sama dengan mean populasi μ, sedangkan varian $\overline{X}$ semakin kecil jika ukuran sampel n bertambah.

Teorema 2 (Teorema Limit Pusat)**.** *Jika $\overline{X}$ adalah mean sampel random berukuran n dari suatu populasi dengan mean μ dan varian σ^2, maka*

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

mendekati berdistribusi normal standar jika n besar.

Umumnya Z mendekati distribusi normal untuk $n \geq 30$. Teorema di atas dapat digunakan untuk mencari nilai pendekatan peluang variabel random $\overline{X}$.

Contoh 3. Suatu perusahan memproduksi bola lampu yang usia hidupnya berdistribusi mendekati normal dengan mean 800 jam dan deviasi standar 40 jam. Berapa peluang suatu sampel random sebanyak 16 bola lampu akan berusia rata-rata kurang dari 775 jam?

Penyelesaian. Distribusi sampel $\overline{X}$ mendekati normal dengan $\mu = 800$ dan $\sigma/\sqrt{n} = 40/\sqrt{16} = 10$. Oleh karena itu peluangnya adalah

$$P(\overline{X} < 775) = P\left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < \frac{775 - 800}{10}\right) = P(Z < -2.5) = 0.0062.$$

Teorema berikut menggambarkan bagaimana sifat nilai harapan dari varian sampel dibandingkan dengan varian populasi.

Teorema 3. *Jika $X_1, X_2, \cdots, X_n$ sampel random berukuran n dari suatu distribusi dengan mean μ dan varian σ^2, maka nilai harapan $S^2 = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{n-1}$ adalah*

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

Dalam inferensi statistik sering diasumsikan bahwa populasinya berdistribusi normal. Hal ini cukup beralasan karena untuk ukuran sampel yang ckup besar sebagaimana dinyatakan dalam Teorema Limit Pusat di atas. Oleh karena itu perlu dibahas distribusi statistik suatu sampel random yang berasal dari populasi berdistribusi normal.

Teorema 4. *Jika $X_1, X_2, \cdots, X_n$ adalah sampel random dengan mean sampel $\overline{X}$, yang diambil dari populasi berdistribusi normal dengan mean μ dan varian σ^2, maka*

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

berdistribusi normal standar.

Teorema 5. *Jika $X_1, X_2, \cdots, X_n$ adalah sampel random dari populasi berdistribusi normal dengan mean μ dan varian σ^2, maka*

$$\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{\sigma^2}$$

berdistribusi chi-square dengan derajat bebas $n-1$.

Bukti. Perhatikan bahwa

$$\begin{aligned}
\sum_{i=1}^{n}(X_i - \mu)^2 &= \sum_{i=1}^{n}[(X_i - \overline{X}) + (\overline{X} - \mu)]^2 \\
&= \sum_{i=1}^{n}(X_i - \overline{X})^2 + \sum_{i=1}^{n}(\overline{X} - \mu)^2 + 2(\overline{X} - \mu)\sum_{i=1}^{n}(X_i - \overline{X}) \\
&= \sum_{i=1}^{n}(X_i - \overline{X})^2 + n(\overline{X} - \mu)^2 \\
&= (n-1)S^2 + \frac{(\overline{X} - \mu)^2}{1/n}.
\end{aligned}$$

Kedua ruang dibagi σ^2 diperoleh

$$\sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} + \frac{(\overline{X} - \mu)^2}{\sigma^2/n}.$$

Karena $\frac{(X_i - \mu)}{\sigma}$ berdistrubusi normal, maka berdasarkan definisi distribusi Chi-square,

$$\sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2}$$

berdistribusi Chi-square dengan derajat bebas n. Demikian pula suku kedua ruang kanan persamaan terkahir yaitu $\frac{(\overline{X} - \mu)^2}{\sigma^2/n}$, berdistrubusi Chi-square dengan derajat bebas 1. Oleh karena itu suku pertama ruang kanan persamaan tersebut, yaitu

$$\frac{(n-1)S^2}{\sigma^2}$$

berdistribusi Chi-square dengan derajat bebas $n-1$.

□

Teorema tersebut dan hasil yang telah diperoleh pada bahasan sebelumnya memberikan hasil berikut.

Teorema 6. *Jika $X_1, X_2, \cdots, X_n$ adalah sampel random dari populasi berdistribusi normal dengan mean μ dan varian σ^2, maka*

$$\frac{\overline{X}-\mu}{S/\sqrt{n}}$$

berdistribusi t dengan derajat bebas $n-1$.

Bukti. Karena $X_1, X_2, \cdots, X_n$ diambil dari populasi berdistribusi normal dengan mean μ dan varian σ^2, maka

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$$

berdistribusi normal standar. Lebih lanjut

$$S^2 = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{n-1}$$

berdistribusi Chi-square dengan derajat bebas n. Oleh karena itu

$$\frac{\sqrt{n}\frac{\overline{X}-\mu}{\sigma}}{\sqrt{S^2/\sigma^2}} = \frac{\overline{X}-\mu}{S/\sqrt{n}}$$

bertistribusi t dengan derajat bebas n. □

Dalam inferensi statistik, sampel random $X_1, X_2, \cdots, X_n$ sering diasumsikan bersifat independen. Jika ukuran populai berhingga, maka tidak ada jaminan sampel random tersebut independen. Namun jika ukuran populasi relatif besar terhadap ukuran sampel, maka sampel random tersebut mendekati independen.

Tugas

1. Hasil pengukuran kadar nikotin pada beberapa jenis rokok diketahui 0.7, 1.2, 2.0, 0.9, 1.1, dan 1.8 miligram. Carilah

   (a) mean sampel,

   (b) varian sampel

2. Suatu perusahaan memproduksi televisi yang usianya berdistribusi normal dengan mean 15 tahun dan deviasi standar 0.5 tahun. Berapa peluang 20 televisi berumur kurang dari 14 tahun?