BAB VII
DISTRIBUSI SAMPLING DAN DESKRIPSI DATA

7.1 Distribusi Sampling (sampling distribution)

Sampling distribution adalah distribusi probabilitas dari suatu statistik. Sampling distribution tergantung dari ukuran populasi, ukuran sampel, metode memililih sampel. Distribusi sampling dari $\overline{X}$ dengan dengan ukuran sampel n adalah suatu distribusi yang bila percobaan dilakukan secara berulang (selalu dengan jumlah sampel n) akan menghasilkan banyak nilai sampel dengan rata-rata $\overline{X}$. Distribusi sampling ini menggambarkan variabilitas (perubahan) rata-rata sampel terhadap rata-rata populasi μ.

7.2 Rata-rata distribusi sampling

Bila suatu sampel acak dari suatu *n* pengamatan diambil dari suatu populasi normal dengan rata-rata μ dan varians σ^2. Maka, setiap pengamatan X_i, i =1,2,3, ..., n dari sampel acak tersebut akan mempunyai distribusi normal yang sama seperti popolasi yang bersangkutan. Sehingga dapat disimpulkan bahwa:

$$\overline{X} = \frac{X_1 + X_2 + \ldots\ldots + X_n}{n}$$; memiliki distribusi normal

dengan rata-rata: $$\mu_{\overline{X}} = \frac{\mu + \mu + \ldots. + \mu}{n} = \mu$$

dan varians: $$\sigma^2_{\overline{X}} = \frac{\sigma^2 + \sigma^2 + \ldots. + \sigma^2}{n^2} = \frac{\sigma^2}{n}$$

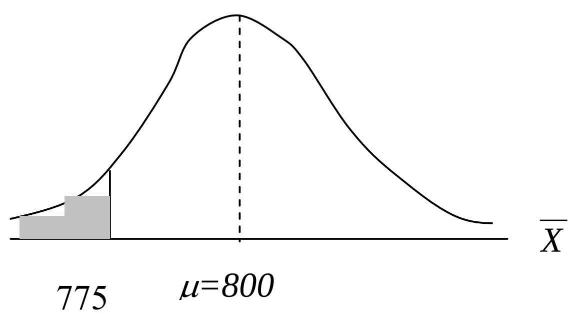
Bila sampel yang diambil dari suatu populasi yang tidak diketahui distribuisnya, distribusi sampling dari $\overline{X}$ akan tetap mendekati nomal dengan rata-rata μ dan varians σ^2 asalkan sampel yang diambil dalam jumlah yang besar. Hasil ini merupakan konsekeuesi dari suatu **teorema batas tengah** *(central limit theorem)* berikut:

Teorema: **Central Limit Theorem**. Bila $\overline{X}$ adalah rata-rata suatu sampel acak yang diambil dari suatu populasi dengan ukuran n, rata-rata μ dan varians σ^2, maka bentuk batas distribusi

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

bila n → ∞, distribusinya adalah distribusi normal standar n(z;0,1).

Contoh: Sebuah perusahaan lampu pijar menyatakan bahwa lampu pijar yang diproduksi mempunyai usia pemakain tersdiribusi normal dengan rata-rata 800 jam dan deviasi standar 40 jam. Tentukan probabilitas bahwa suau sampel acak dari 16 buah lampu pijar tersebut memiliki rata-rata usia kurang dari 775 jam.



Solusi: Distribusi sampling $\overline{X}$ akan mendekati normal dengan $\mu_{\overline{X}} = 800$ dan $\sigma_{\overline{X}} = 40 / \sqrt{16} = 10$; probabilitas yang diinginkan seperti gambar diatas. Maka:

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} = \frac{775 - 800}{10} = -2.5$$

$$P\left(\overline{X} < 775\right) = P(Z < -2.5) = 0.0062$$

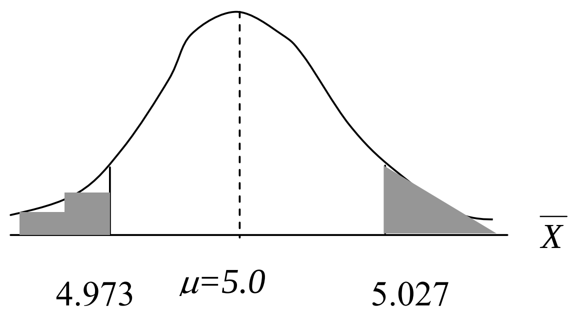
__Inferensi Rata-rata populasi__

Salah satu aplikasi penting dari teorema batas tengah adalah penentuan nilai rata-rata populasi yang pantas.

Contoh: Sebuah industri memproduksi suatu komponen alat suku cadang mobil. Hal yang penting dalam proses produksi komponen tersebut adalah rata-rata diameter ukuran komponen 5 mm. Para teknisi yang terlibat dalam pembuatan kompenen tersebut memperkirakan rata-rata diameter populasi adalah 5.0 mm. Sebuah percobaan dilukakan pada 100 komponen yang dipilih secara acak dan masing-masing komponen diukur. Diketahui bahwa deviasi standar σ = 0.1. Hasil percobaan menunjukkan rata-rata diameter sampel $\overline{X}$ = 5.027 mm. Apakah informasi sampel ini mendukung perkiraan teknisi tersebut?

Solusi:

$$P\left[\left|\left(\overline{X}-5\right)\right| \geq 0.027\right]$$



Bila rata-rata μ = 5, berapa peluang bahwa $\overline{X}$ akan menyimpang sebesar 0.027 mm?

$$P\left[\left|\left(\overline{X}-5\right)\right| \geq 0.027\right] = P\left[\left(\overline{X}-5\right) \geq 0.027\right] + P\left[\left(\overline{X}-5\right) \leq -0.027\right]$$

$$= 2P\left(\frac{\overline{X}-5}{0.1/\sqrt{100}} \geq 2.7\right)$$

Bila perkiraan μ = 5.0 adalah benar maka $\dfrac{\overline{X}-5.0}{0.1/\sqrt{100}}\, adalah\ N(0,1)$.

$$2P\left(\frac{\overline{X}-5}{0.1/\sqrt{100}} \geq 2.7\right) = 2P\left[Z \geq 2.7\right] = 2(0.0035) = 0.007$$

Hal ini menunjukkan bahwa penyimpangan 0.027 hanya terjadi pada 7 komponen dalam 1000 percobaan. Sehingga percobaan diatas dengan $\overline{X}$ = 5.027 mm tidak mendukung perkiraan teknisi bahwa μ = 5.0.

Distribusi sampling dari dua rata-rata

Teorema: Bila dua sampel yang saling bebas, n_1 dan n_2, diambil dari dua populasi, diskrit atau menerus, dengan rata-rata μ_1 dan μ_2 dan varians σ_1^2 dan σ_2^2, maka distribusi sampling dari perbedaan rata-rata, $\overline{X}_1$ - $\overline{X}_2$ mendekati distribusi normal dengan rata-rata dan varians sebagai berikut:

$$\mu_{\overline{X}_1 - \overline{X}_2} = \mu_1 - \mu_2 \quad \text{dan} \quad \sigma^2_{\overline{X}_1 - \overline{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Sehingga:

$$Z = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\left(\sigma_1^2 / n_1\right) + \left(\sigma_2^2 / n_2\right)}} \rightarrow \text{mendekati variabel normal standar}$$

Contoh: Tabung televisi dari perusahaan A dan B mempunyai rata-rata *lifetime*, deviasi standar dan jumlah sampel yang diambil sebagai berikut:

Perusahaan A	Perusahaan B
$\mu_A = 6.5$	$\mu_B = 6.0$
$\sigma_A = 0.9$	$\sigma_B = 0.8$
$n_A = 36$	$n_B = 49$

Berapa probabilitas bahwa suatu sampel acak dari 36 tabung perusahaan A akan mempunyai rata-rata umur hidup paling tidak 1 tahun lebih lama dari rata-rata umur hidup 49 tabung perusahaan B.

Solusi:

$$\mu_{\overline{X}_A - \overline{X}_B} = \mu_A - \mu_B = 6.5 - 6.0 = 0.5$$

$$\sigma_{X_A - X_B} = \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_A^2}{n_B}} = \sqrt{\frac{0.81}{36} + \frac{0.64}{49}} = 0.189$$

$\overline{X}_A$ - $\overline{X}_B$=1

$$Z = \frac{1.0 - 0.5}{0.189} = 2.65$$

$$P\left(\overline{X}_A - \overline{X}_B \geq 1.0\right) = P(Z > 2.65) = 1 - P(Z < 2.65) = 1 - 0.9960 = 0.0040$$

7.3 Distribusi sampling S^2

Bila S^2 adalah varians dari suatu *n* sampel acak yang diambil populasi normal dengan nilai varians σ^2 , maka statitsik:

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

Memiliki distribusi *chi-squared* dengan derajat kebebasan $v = n$ -1.

Contoh: Pabrik baterei menjamin bahwa rata-rata umur batereinya 3 tahun dan deviasi standar 1 tahun. Bila 5 dari sampel baterei yang diambil umur baterinya adalah 1,9; 2,4; 3,0; 3,5 dan 4,2 tahun. Apakah produsen baterei tersebut masih tetap yakin bahwa standar deviasinya 1 tahun. Asumsikan umur baterei memiliki distribusi *chi-squared*.

Solusi:

Varians dari *n* sampel acak adalah:

$$S^2 = \frac{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}{n(n-1)} = \frac{(5)(48.26) - (15)^2}{(5)(4)} = 0.815$$

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{(4)(0.815)}{1} = 3.26$$

Karena 95% dari nilai χ^2 dengan derajat kebebasan 4 terletak antara 0.711 dan 9.488; maka nilai yang dihitung dengan σ^2 masih layak (produsen tetap yakin bahwa standar deviasinya 1 tahun).

7.4 t-Distribution

Teorema: Misalkan Z adalah variabel acak normal standar dan V adalah variabel acak distribusi *chi-squared* dengan derajat kebebasan v. Bila Z dan V adalah independent, maka distribusi variabel acak T,

$$T = \frac{Z}{\sqrt{V/v}}$$

dengan fungsi kepadatan;

$$h(t) = \frac{\Gamma[(v+1)/2]}{\Gamma(v/2)\sqrt{\pi v}}\left(1+\frac{t^2}{v}\right)^{-(v+1)/2} \quad ; \quad -\infty < t < \infty$$

Disebut sebagai *t-distribution* dengan derajat kebebasan v.

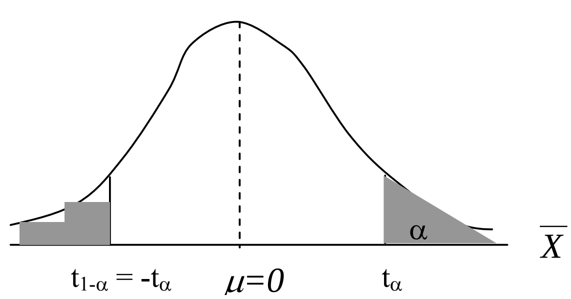
Corollary: Misalkan $X_1, X_2, ..., X_n$ adalah variabel acak bebas yang semuanya normal dengan rata-rata μ dan standar deviasi σ. Bila:

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

Maka variabel acak $T = \dfrac{\overline{X} - \mu}{s/\sqrt{n}}$ memiliki distribusi-t dengan derajat kebebasan $v = n - 1$.

t-distribution sering disebut juga dengan *Student t-distribution*.

Distribusi t mirip dengan distribusi normal yakni keduanya simteris dan rata-rata = 0. Karena distribusi-t simetris maka: $t_{1-\alpha} = -t_\alpha$



Contoh: Tentukan nilai-t dengan derajat kebebasan 14 dan luas area α=0.025 disebelah kiri, atau 0.975 disisi kanan.

Solusi: lihat tabel distribusi t:

$$T_{0.975} = -t_{0.025} = -2.145$$

Contoh: Tentukan $P(-t_{0.025} < T < t_{0.05})$

Solusi: Karena luas area disisi kanan = 0.05 dan sebelah kiri 0.025; maka;

Luas area T = 1- 0.05 – 0.025 = 0.925, sehingga

$P(-t_{0.025} < T < t_{0.05}) = 0.925$

Contoh: Seorang insinyur kimia menyatakan bahwa rata-rata populasi yang dihasilkan dari suatu proses produksi adalah 500 grams/mm. Untuk membuktikan pernyataan tersebut diambil 25 sampel. Bila nilai t berada antara $-t_{0.05}$ dan $t_{0.05}$ maka pernyataan tersebut benar. Dari hasil percobaan terhadap 25 sampel tersebut menunjukkan bahwa rata-rata $\bar{x}$ =518 gram/mm dan deviasi standar s = 40 gram.

Solusi: dari tabel menunjukkan bahwa untuk derajat kebebasan 24 $t_{0.05}$ = 1.711. Oleh karena itu, pernyataan tsb benar bila nilai t antara -1.711 dan 1.711. bila μ = 500, maka:

$$t = \frac{\overline{X} - \mu}{s/\sqrt{n}} = \frac{518 - 500}{40/\sqrt{25}} = 2.25$$

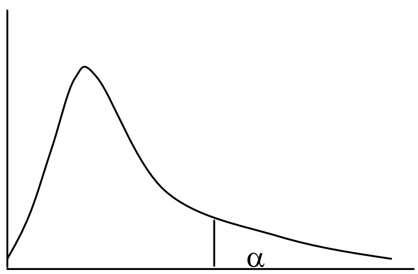
Probabilitas nilai t sama atau lebih besar dari 2.25 dengan v = 24 adalah mendekati 0.02. Bila μ > 500, menghasilkan nilai t yang lebih pantas. Oleh karena itu, hasil produksinya lebih baik dari yang mereka nyatakan.

7.5. F-distribution

Teorema: Misalkan U dan V adalah variabel acak bebas dengan distribusi *chi-squared* dan v_1 & v_2 adalah deajat kebebasan. Maka distribusi variabel acak $F = \dfrac{U/v_1}{V/v_2}$ dengan fungsi kepadatan sbb:

$$h(f) = \begin{cases} \dfrac{\Gamma[(v_1+v_2)/2](v_1/v_2)^{v_1/2}}{\Gamma(v_1/2)\Gamma(v_2/2)} \cdot \dfrac{f^{v_1/2-1}}{(1+v_1 f/v_2)^{(v_1+v_2)/2}} & 0 < f < \infty \\ 0 & \text{selainnya} \end{cases}$$

adalah **F-distribution.**



Teorema:

$$f_{1-\alpha}(v_1, v_2) = \frac{1}{f_\alpha(v_2, v_1)}$$

Sehingga nilai f dengan 6 dan 10 derajat kebebasan berada pada daerah sisi kanan 0.95 adalah: (lihat tabel distribusi F)

$$f_{0.95}(6,10) = \frac{1}{f_{0.05}(10,6)} = \frac{1}{4.06} = 0.246$$

Teorema: Bila S_1^2 dan S_2^2 adalah varians dari n_1 dan n_2 sampel acak bebas yang diambil dari populasi normal dengan varians σ_1^2 dan σ_2^2, maka:

$$F = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}$$ adalah Distribusi F

dengan derajat kebebasan $v_1 = n_1 - 1$ dan $v_2 = n_2 - 1$