DISTRIBUSI SAMPLING

Definisi : distribusi sampling adalah distribusi peluang untuk nilai statistik yang diperoleh dari sampel acak untuk menggambarkan populasi.

1. Distribusi rata-rata

Misal sampel acak n diambil dari populasi normal dengan rataan μ dan varians σ^2. Tiap pengamatan X_i, i = 1, 2, 3, …, n dari sampel acak tersebut akan berdistribusi normal yang sama dengan populasi yang diambil sampelnya. Jadi

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Berdistribusi normal dengan rataan

$$\mu_{\bar{x}} = \frac{\mu + \mu + \cdots + \mu}{n}$$

Dan variasi

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2 + \sigma^2 + \cdots + \sigma^2}{n^2}$$

Bila populasi yang disampel tidak diketahui distribusinya, berhingga atau tidak, maka distribusi sampel $\bar{X}$ masih akan berdistribusi hampir normal dengan rataan μ dan varians σ^2/n asalkan ukuran sampelnya besar. Ini merupakan akibat dari **Teorema Limit Pusat.**

Teorema Limit Pusat

Bila $\bar{X}$ rataan sampel acak ukuran n yang diambil dari populasi dengan rataan μ dan varians σ^2 yang berhingga, maka bentuk limit dari distribusi

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

Bila $n \to \infty$, ialah distribusi normal baku $n(z; 0,1)$

Hampiran normal untuk $\bar{X}$ umumnya cukup baik bila $n \geq 30$, terlepas dari bentuk populasi. Bila $n < 30$, hampiran hanya akan baik bila populasinya tidak jauh berbeda dengan normal. Bila populasi diketahui normal, maka distribusi sampel $\bar{X}$ akan tepat berdistribusi normal, dan ukuran sampelnya tidak menjadi soal.

Untuk $n < 30$ atau $\frac{n}{N} > 5\%$ maka

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Berdistribusi normal dengan rataan

$$\mu_{\bar{x}} = \frac{\mu + \mu + \cdots + \mu}{n}$$

Dan variasi

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2 + \sigma^2 + \cdots + \sigma^2}{n^2} . \frac{N - n}{N - 1}$$

Contoh:

Tinggi badan mahasiswa rata-rata mencapai 165 cm dan simpangan baku 8,4cm. Telah diambil sebuah sampel acak terdiri atas 45 mahasiswa. Tentukan berapa peluang tinggi rata-rata ke 45 mahasiswa tersebut:

a. Antara 160 cm dan 168 cm
b. Paling sedikit 166 cm

Jawab

Jika ukuran populasi tidak disebutkan besarnya selalu dianggap cukup besar.

$\mu_{\bar{x}} = 165\ cm$ dan $\sigma_{\bar{x}}^2 = \frac{8,4}{\sqrt{45}} = 1,252$

a. $P(160 < \bar{X} < 168) = P\left(\frac{160-165}{1,252} < Z < \frac{168-165}{1,252}\right)$
$= P(-3,99 < Z < 2,4)$

Dengan menggunakan tabel distribusi normal (sudjana, 2005:490) diperoleh
= 0,5 + 0,4918 = 0,9918

b. $P(\bar{X} \geq 166) = P\left(Z \geq \frac{166-165}{1,252}\right) = P(Z \geq 0,80) = 0,5 - P(Z < 0,80) = 0,5 - 0,2881 = 0,2119$

2. Distribusi selisih dan jumlah rata-rata

Misalkan ada dua populasi, yang pertama dengan rataan μ_1 dan varians σ_1^2, yang kedua dengan rataan μ_2 dan varians σ_2^2. Misalkan statistik $\overline{X_1}$ menyatakan rataan sampel acak ukuran n_1 yang diambil dari populasi pertama, dan statistik $\overline{X_2}$ menyatakan rataan sampel acak ukuran n_2 yang diambil dari populasi kedua; kedua sampel saling bebas satu sama lain.

SELISIH

Peubah $\overline{X_1}$ dan $\overline{X_2}$ keduanya berdistribusi hampir normal masing-masing dengan rataan μ_1 dan μ_2, dan varians σ_1^2/n dan σ_2^2/n. Hampiran ini bertambah baik bila n_1 dan n_2 membesar maka $\overline{X_1} - \overline{X_2}$ berdistribusi normal dengan rataan

$$\mu_{\overline{x_1}-\overline{x_2}} = \mu_{\overline{x_1}} - \mu_{\overline{x_2}} = \mu_1 - \mu_2$$

Dan variansi

$$\sigma_{\overline{x_1}-\overline{x_2}}^2 = \sigma_{\overline{x_1}}^2 + \sigma_{\overline{x_2}}^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Teorema

Bila sampel bebas ukuran n_1 dan n_2 diambil secara acak dari dua populasi, masing-masing dengan rataan μ_1 dan μ_2, dan varians σ_1^2 dan σ_2^2, maka distribusi sampel dari selisih rataan $\overline{X_1} - \overline{X_2}$, berdistribusi hampir normal rataan dan variansi diberikan dengan

$$\mu_{\overline{x_1}-\overline{x_2}} = \mu_1 - \mu_2 \text{ dan } \sigma_{\overline{x_1}-\overline{x_2}}^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

sehingga

$$Z = \frac{(\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Secara hampiran merupakan peubah normal baku

Jika n_1 dan n_2 keduanya lebih besar sama dengan 30, hampiran normal untuk distribusi $\overline{X_1} - \overline{X_2}$ sangat baik tidak tergantung dari bentuk kedua populasi. Akan tetapi, jika n_1 dan n_2 kurang dari 30, hampiran normal lumayan baik kecuali bila kedua populasi agak jauh dari normal. Tentu saja bila kedua populasi normal, maka $\overline{X_1} - \overline{X_2}$ berdistribusi normal terlepas dari ukuran n_1 dan n_2.

JUMLAH

Peubah $\overline{X_1}$ dan $\overline{X_2}$ keduanya berdistribusi hampir normal masing-masing dengan rataan μ_1 dan μ_2, dan varians σ_1^2/n dan σ_2^2/n. Hampiran ini bertambah baik bila n_1 dan n_2 membesar maka $\overline{X_1} + \overline{X_2}$ berdistribusi normal dengan rataan

$$\mu_{\overline{x_1}+\overline{x_2}} = \mu_{\overline{x_1}} + \mu_{\overline{x_2}} = \mu_1 + \mu_2$$

Dan variansi

$$\sigma^2_{\overline{x_1}+\overline{x_2}} = \sigma^2_{\overline{x_1}} + \sigma^2_{\overline{x_2}} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Teorema

Bila sampel bebas ukuran n_1 dan n_2 diambil secara acak dari dua populasi, masing-masing dengan rataan μ_1 dan μ_2, dan varians σ_1^2 dan σ_2^2, maka distribusi sampel dari jumlah rataan $\overline{X_1} + \overline{X_2}$, berdistribusi hampir normal rataan dan variansi diberikan dengan

$$\mu_{\overline{x_1}+\overline{x_2}} = \mu_1 + \mu_2 \text{ dan } \sigma^2_{\overline{x_1}+\overline{x_2}} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

sehingga

$$Z = \frac{(\overline{X_1} + \overline{X_2}) - (\mu_1 + \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Secara hampiran merupakan peubah normal baku

Jika n_1 dan n_2 keduanya lebih besar sama dengan 30, hampiran normal untuk distribusi $\overline{X_1} + \overline{X_2}$ sangat baik tidak tergantung dari bentuk kedua populasi. Akan tetapi, jika n_1 dan n_2 kurang dari 30, hampiran normal lumayan baik kecuali bila kedua populasi agak jauh dari normal. Tentu saja bila kedua populasi normal, maka $\overline{X_1} + \overline{X_2}$ berdistribusi normal terlepas dari ukuran n_1 dan n_2.

Contoh:

Suatu sampel berukuran $n_1 = 140$ diambil secara acak dari populasi yang berdistribusi normal dengan rataan $\mu_1 = 163$ dan simpangan baku $\sigma_1 = 5{,}2$ dan rataan sampel $\overline{x_1}$ dihitung. Sampel acak berukuran $n_2 = 140$ diambil, bebas dari yang pertama, dari populasi lain yang juga berdistribusi normal, dengan rataan $\mu_2 = 152$ dan simpangan baku $\sigma_2 = 4{,}9$ dan rataan sampel $\overline{x_2}$ dihitung. Tentukan peluang rataan sampel pertama paling sedikit lebih 10 dari rataan sampel kedua.

Jawab:

Dari distribusi sampel $\overline{X_1} - \overline{X_2}$ diketahui bahwa distribusinya normal dengan rataan

$$\mu_{\overline{x_1}-\overline{x_2}} = \mu_1 - \mu_2 = 163 - 152 = 11$$

Dan simpangan baku

$$\sigma^2_{\overline{x_1}-\overline{x_2}} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \frac{5{,}2^2}{140} + \frac{4{,}9^2}{140} = 0.3646 \leftrightarrow \sigma_{\overline{x_1}-\overline{x_2}} = 0{,}6038$$

$$P(\overline{x_1} - \overline{x_2} \geq 10) = P\left(Z \geq \frac{10-11}{0{,}6038}\right) = P(Z \geq -1{,}66) = 0{,}5 + 0{,}4515 = 0{,}9515$$

Contoh:

Rata-rata tinggi mahasiswa laki-laki 163 cm dan simpangan bakunya 5,2 cm; sedangkan untuk mahasiswa perempuan, parameter tersebut berturut-turut 152 cm dan 4,9 cm.

Dari kedua kelompok mahasiswa itu masing-masing diambil sebuah sampel acak, secara independen, berukuran sama, ialah 140 orang. Berapa peluang rata-rata tinggi mahasiswa laki-laki paling sedikit 10 cm lebihnya dari rata-rata tinggi mahasiswa perempuan?

Jawab:
Misalkan $\bar{x}$ dan $\bar{y}$ masing-masing menyatakan rata-rata tinggi dari sampel untuk mahasiswa laki-laki dan perempuan. Yang ditanyakan adalah $P(\bar{x} - \bar{y} \geq 10)$.
Diketahui: $\mu_1 = \mu_x = 163\ cm$, $\mu_2 = \mu_y = 152\ cm$, $\sigma_1 = \sigma_x = 5.2$, $\sigma_2 = \sigma_y = 4.9$ dan $n_1 = n_2 = 140$. Menurut teori diatas, $\bar{x} - \bar{y}$ berdistribusi normal dengan rata-rata $\mu_{\bar{x}-\bar{y}} = (163 - 152) = 11$ dan simpangan baku $\sigma_{\bar{x}-\bar{y}} = \sqrt{\frac{(5,2)^2}{140} + \frac{(4,9)^2}{140}} = 0.6038$
Maka diperoleh
$$P\left(Z \geq \frac{10-11}{0.6038}\right) = P(Z \geq -1.66) = 0.5 + P(Z < 1.66) = 0.5 + 0.4515 = 0.9515$$

3. Distribusi proporsi
Misalkan populasi berukuran N yang didalamnya terdapat kejadian A sebanyak Y. Maka proporsi kejadian A sebesar $\pi = \frac{Y}{N}$.
Dari populasi ini diambil sampel acak berukuran n dan dimisalkan didalamnya ada peristiwa A sebanyak x. Sampel ini memberikan statistik proporsi peristiwa A = $\frac{x}{n}$. Maka $\frac{x}{n}$ berdistribusi normal dengan rataan
$$\mu_{x/n} = \pi$$
Dan variansnya
$$\sigma^2_{x/n} = \frac{\pi(1-\pi)}{n}$$

> Untuk ukuran sampel n cukup besar, berlaku sifat:
> Jika dari populasi yang berdistribusi binom dengan parameter π untuk peristiwa A, $0 < \pi < 1$, diambil sampel acak berukuran n dimana statistik proporsi untuk peristiwa A = $\frac{x}{n}$, maka untuk n cukup besar, distribusi proporsi $\frac{x}{n}$ mendekati distribusi normal dengan parameter $\mu_{x/n} = \pi$ dan variansnya $\sigma^2_{x/n} = \frac{\pi(1-\pi)}{n}$, sehingga
> $$Z = \frac{\frac{x}{n} - \pi}{\sigma_{x/n}}$$
> Secara hampiran merupakan peubah normal baku

Untuk $n < 30$ atau $\frac{n}{N} > 5\%$ maka $\frac{x}{n}$ berdistribusi normal dengan rataan
$$\mu_{x/n} = \pi$$
Dan variasi
$$\sigma^2_{x/n} = \frac{\pi(1-\pi)}{n} . \frac{N-n}{N-1}$$

Contoh:
Apa petunjuk kuat bahwa 10% anggota masyarakat tergolong ke dalam golongan A. Sebuah sampel acak terdiri atas 100 orang telah diambil. Tentukan peluangnya bahwa dari 100 orang itu akan ada paling sedikit 15 orang dari golongan A.
Jawab:
Populasi yang dihadapi berukuran cukup besar dengan $\pi = 0.1$ dan $1 - \pi = 0.9$
Untuk ukuran sampel 100, diantaranya paling sedikit 15 tergolong kategori A, maka paling sedikit $\frac{x}{n} = 0.15$.

Kekeliruan bakunya adalah:

$$\sigma_{x/n} = \sqrt{\frac{(0.1)(0.9)}{100}} = 0.03$$

Maka peluangnya

$$P\left(Z \geq \frac{0.15 - 0.1}{0.03}\right) = P(Z \geq 1.67) = 0.5 - P(Z < 1.67) = 0.5 - 0.4525 = 0.0475$$

4. Distribusi selisih proporsi

Misalkan ada dua populasi masing-masing berdistribusi binom, kedua-duanya berukuran cukup besar. Didalam kedua populasi itu ada peristiwa A dengan proporsi masing-masing populasi secara berturut-turut yaitu π_1 dan π_2. Dari kedua populasi diambil sampel acak secara independen, sebanyak n_1 dari populasi satu dan sebanyak n_2 dari populasi dua. Untuk peristiwa A, didapat kumpulan proporsi

$$\frac{x_i}{n_1}, i = 1,2, \dots, k \text{ dan } \frac{y_j}{n_2}, j = 1,2, \dots, r$$

dengan x_i = banyak peristiwa A dalam sampel yang diambil dari populasi satu, y_j = banyak peristiwa A dalam sampel yang diambil dari populasi dua, k dan r masing-masing banyak sampel yang mungkin diambil dari populasi kesatu dan populasi kedua.

Selisih proporsi $\left(\frac{x_i}{n_1} - \frac{y_j}{n_2}\right)$ dapat dibentuk sehingga terdapat kumpulan selisih proporsi. Dari kumpulan ini dapat dihitung rata-ratanya, diberi simbol μ_{sp} dan simpangan baku, siberi simbol σ_{sp}, dengan $sp = \left(\frac{x_i}{n_1} - \frac{y_j}{n_2}\right)$ = selisih antara proporsi sampel kesatu dan proporsi sampel kedua. Rata-rata dan simpangan baku tersebut juga dapat dihitung dengan formula:

$$\mu_{sp} = \pi_1 - \pi_2$$

$$\sigma_{sp} = \sqrt{\frac{\pi_1(1 - \pi_1)}{n_1} + \frac{\pi_2(1 - \pi_2)}{n_2}}$$

Untuk ukuran-ukuran sampel n_1 dan n_2 cukup besar, biasanya $n_1 \geq 30$ dan $n_2 \geq 30$, maka distribusi selisih proporsi ini akan mendekati distribusi normal dengan parameter- parameter $\mu_{sp} = \pi_1 - \pi_2$ dan $\sigma_{sp} = \sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}$. Agar supaya distribusi normal ini menjadi distribusi normal baku maka diperlukan transformasi.

$$Z = \frac{\left(\frac{x}{n_1} - \frac{y}{n_2}\right) - (\pi_1 - \pi_2)}{\sigma_{sp}}$$

Contoh:

Ada petunjuk kuat bahwa calon A akan mendapat suara 60% dalam pemilihan. Dua buah sampel acak secara independen telah diambil masing-masing terdiri atas 300 orang. Tentukan peluangnya akan terjadi perbedaan persentase tidak lebih dari 10% yang akan memilih A.

Jawab:

Kedua sampel diambil dari sebuah populasi, jadi dianggap kedua populasi yang sama, sehingga $\pi_1 = \pi_2 = 0.6$. Jika x = banyak orang yang memilih A dalam sampel kedua, dan y = banyak orang yang memilih A dalam sampel kedua, maka yang dicari adalah peluang $\left(\frac{x}{n_1} - \frac{y}{n_2}\right) < 0.1$ atau $\left(\frac{y}{n_2} - \frac{x}{n_1}\right) < 0.1$

Setelah digabungkan menjadi $-0.1 < \left(\frac{x}{n_1} - \frac{y}{n_2}\right) < 0.1$

Maka

$$\mu_{sp} = 0.6 - 0.6 = 0$$

Dan

$$\sigma_{sp} = \sqrt{\frac{(0.6)(0.4)}{300} + \frac{(0.6)(0.4)}{300}} = 0.04$$

Sehingga

$$P\left(-0.1 < \left(\frac{x}{n_1} - \frac{y}{n_2}\right) < 0.1\right) = P\left(\frac{-0.1 - 0}{0.04} < Z < \frac{0.1 - 0}{0.04}\right) = P(-2.5 < Z < 2.5)$$
$$= 2P(Z < 2.5) = 2(0.4938) = 0.9876$$

Daftar Pustaka

Mendenhall, W., Beaver, R., Beaver, B. 2006. *Introduction to Probability and Statistics*. USA: Thomson Brooks/Cole

Sudjana. 2005. *Metode Statistika*. Bandung: Tarsito

Walpole, R., Myers, R. 1995. *Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuwan*. Bandung: ITB