DISTRIBUSI SAMPLING

MODUL PERKULIAHAN 12 (ONLINE 10)

Disusun oleh:

TIM DOSEN

Pelaksana Akademik Mata Kuliah Umum (PAMU)

Universitas Esa Unggul

Jakarta Barat

2018

DISTRIBUSI SAMPLING

Tujuan mempelajari distribusi sampling adalah sebagai berikut.

- Memahami perlunya suatu sampling (pengambilan sampel) serta keuntungan-keuntungan melakukannya
- Menjelaskan pengertian sampel acak untuk sampling tanpa pergantian untuk suatu populasi terhingga dan pengambilan sampel untuk populasi tak terhingga
- Menjelaskan langkah-langkah yang diperlukan untuk membentuk suatu distribusi sampling dari mean-mean sampel, menghitung mean dan deviasi standard dari distribusi sampling tersebut
- Menjelaskan langkah-langkah yang diperlukan untuk membentuk suatu distribusi sampling dari proporsi sampel, menghitung mean dan deviasi standard dari distribusi sampling tersebut
- Menghitung mean dan deviasi standard dari distribusi sampling yang merupakan perbedaan atau penjumlahan dari sampel-sampel yang berasal dari dua populasi

Pokok bahasan dalam modul ini adalah sebagai berikut.

- Pengertian dan Konsep Dasar Sampling
- Distribusi Sampling Dari Mean
- Distribusi Sampling Dari Proporsi
- Distribusi Sampling Dari Perbedaan dan Penjumlahan

Contohnya, sebuah lembaga survey melakukan polling menjelang hari pemilu di Indonesia (mengambil sekitar 1.500 s/d 2.000 pemilih sebagai sampel untuk diteliti). Selanjutnya, hasil analisis terhadap sampel tersebut digunakan untuk menduga populasi. Misalnya dari sampel diketahui 45% memilih Prabowo, dan 55% memilih Jokowi, maka kedua angka tersebut merupakan penduga untuk proporsi populasi pemilih di Indonesia yang memilih Prabowo dan Jokowi. Bukti empiris menunjukkan bahwa dari 13 kali pemilihan presiden di Indonesia, hanya sekali lembaga survey membuat kesalahan prediksi.

Kebutuhan dan keuntungan sampling

Sampling yang baik adalah sampling yang dapat menghemat biaya biaya dan waktu, serta menjaga keakuratan hasil-hasilnya. Secara khusus teknik sampling berguna dalam Estimasi parameter populasi (seperti mean populasi, varians populasi dll.) yang tidak diketahui berdasarkan pengetahuan tentang statistik sampel (seperti mean sampel, varians sampel, dll.) yang berkaitan, menentukan apakah perbedaan yang teramati pada dua sampel adalah benar-benar signifikan (berarti) atau karena variasi yang kebetulan sifatnya.

Metode Penarikan Sampel

1. Penarikan sampel probabilitas:

   - prosedur objektif: probabilitas pemilihan diketahui terlebih dahulu untuk setiap elemen populasi.
   - setiap elemen populasi memiliki probabilitas yang sama sebagai sampel.
   - metode pemilihan acak (random), konsep matematik yang tepat , sehingga setiap elemen dalam populasi memiliki peluang yang sama sebagai sampel.

2. Penarikan sampel non probabilitas:

   - prosedur subjektif, kerangka sampelnya tidak tersedia.
   - Setiap elemen populasi tidak memiliki probabilitas yang sama sebagai sampel, dipilih berdasarkan pertimbangan-pertimbangan pribadi.

PROBABILITY SAMPLING

1. **Sampling acak sederhana (*simple random sampling*)**

- Baik (bukti empiris yang dihasilkan), representatif
- Populasi terbatas: peluang acak secara individual.
- Populasi banyak dan berkelompok: mengambil sejumlah kelompok yang ada, kemudian pengambilan sampel acak dilakukan pada kelompok tersebut.
  Misalnya, sampel = 35 secara acak dari populasi=100, (*dealer* sepeda motor X di Jakarta, Bandung dan Surabaya). Masing-masing nama dealer diberi nomor sampai dengan 100, kemudian setiap nomor ditulis pada secarik kertas dan selanjutnya kertas-kertas bernomor tersebut dimasukkan ke dalam sebuah kotak, lalu dikocok dengan baik, selanjutnya dipilih sebanyak 35 sampel yang prosedur penarikannya dilakukan 35 kali. Cara lain adalah dengan menggunakan tabel bilangan acak.

tabel bilangan acak

27104	54374	83559	75559	90159	12338	36248	65036	47393	26242
85694	38272	69261	97632	94143	55827	37871	82946	18894	19132
77853	22702	28785	51676	02936	82362	73695	41692	19725	15049
86600	07664	89694	99067	51985	30784	95287	18448	91182	29128
26383	75981	54705	99067	93115	50695	57523	11214	64728	88875
45827	58215	58603	43798	31221	78476	78063	44014	67083	14321
23011	53152	15022	23592	89899	37661	17709	99827	73371	18303
28398	55932	20704	73139	96574	23366	21128	21770	43886	23808
99848	26278	14123	74472	97826	09121	00773	06158	65603	65568
62758	51058	48298	60557	72308	62076	05105	78524	39564	10347
98136	36786	33878	13646	72354	54715	19117	18929	54414	14321
18677	97088	89968	78156	26378	51126	83467	98723	85121	18308
40571	59619	34135	69444	45123	89107	15229	49271	12864	23808
62454	42436	94025	42152	52746	09096	53333	50151	48104	65568
55317	75637	02969	78351	05005	29496	95398	26237	13411	10347
04328	16840	16840	28042	94563	91678	04528	49891	34409	99024
17155	50891	53891	81421	82014	89254	44901	69374	75771	82675
50979	77093	36204	69750	52800	51098	94739	15749	43529	45405
31476	71195	59846	77236	21867	84452	32977	87833	51947	20388
28663	01447	60287	80016	35196	50923	30967	90195	18865	59892

b. **Sampling acak berstrata disproporsional**

- Bila populasi berstrata, tetapi kurang proporsional. (kasus di atas, secara disproporsional dapat ditarik sampel, misalnya untuk wanita 60% = 60 dan pria 40% = 40).

- Prinsip sampling disproporsional adalah :

  – Semakin besar suatu strata, semakin besar sampel
  – Semakin tinggi variabilitas di dalam suatu sampel, semakin besar sampel

Misalnya, pegawai dari unit kerja tertentu mempunyai 3 orang lulusan S3, 4 orang lulusan S2, 90 orang lulusan S1, 800 orang lulusan SMU dan 700 orang lulusan SMP. Dalam hal ini, 3 orang lulusan S3 dan 4 orang lulusan S2 diambil semuanya sebagai sampel, karena dua kelompok ini terlalu kecil bila dibandingkan dengan kelompok S1, SMU dan SMP.

2. **sampling acak berstrata proporsional (*proportioned stratified random sampling*)**

   Subsample-subsampel acak sederhana ditarik dari setiap strata yang kurang lebih sama dalam beberapa karakteristik.

   a. **Sampling acak berstrata proporsional**

   Bila populasi mempunyai anggota/unsur tidak homogen dan berstrata secara proporsional. Untuk suatu organisasi yang mempunyai pegawai dengan latar belakang pendidikan berstrata, populasi pegawai itu berstrata. Misalnya, populasi = 1000 (700 orang wanita dan 300 orang pria). Sampel yang diperlukan = 100. Secara proporsional, sampel yang dapat ditarik adalah wanita = 700/1000 * 100 = 70 dan pria = 300/1000 * 100 = 30.

3. **Metode sampling berkelompok (*cluster sampling*)**

   - Memilih subpopulasi yang disebut klaster, setiap elemen kelompok dipilih sebagai anggota sampel.
   - Untuk objek dengan data sangat luas (penduduk Negara, provinsi) samplingnya berdasarkan daerah populasi yang telah ditetapkan.
   - Kriteria *cluster* bertolak belakang dengan apa yang digunakan dalam sampling berstrata.
   - Populasi harus dibagi ke dalam kelompok-kelompok yang bersifat *mutually exclusive, s*elanjutnya dipilih secara acak sebagai sampel.
     Misal, populasi (20 elemen, 4 kelompok ukuran sama)

Kelompok	Jumlah elemen populasi
1	1, 2, 3, 4, 5
2	6, 7, 8, 9, 10
3	11, 12, 13, 14, 15
4	16, 17, 18, 19, 20

Lalu dipilih secara acak kelompok-kelompok yang akan dijadikan sampel. Kemudian, dari kelompok yang terpilih, anggota-anggota kelompok tersebut dipilih secara acak untuk dijadikan sampel.

Contoh lain, Indonesia terdiri dari 30 provinsi dan sampelnya akan menggunakan 15 provinsi. Pengambilan 15 provinsi tersebut dilakukan secara acak. Tetapi karena provinsi di Indonesia adalah berstrata (tidak sama), sehingga perlu menggunakan sampling acak berstrata. Ada provinsi di Indonesia yang penduduknya padat, ada yang tidak, ada yang mempunyai hutan banyak, ada yang tidak, ada yang kaya bahan tambang, dan ada yang tidak. Karakteristik semacam ini perlu diperhatikan sehingga pengambilan sampel menurut strata populasi dapat ditetapkan.

NONPROBABILITY SAMPLING

- Prosedur bersifat subjektif.
- Probabilitas pemilihan elemen populasi tidak dapat ditentukan.
- Hemat waktu/biaya (tidak perlu kerangka sampling)
- Hasilnya bisa bias dan ketidakpastian.

- Misalnya, dalam suatu penelitian terhadap para pengunjung mal atau pusat-pusat perbelanjaan.

a. **Sampling Sistematik**

Berdasarkan urutan anggota populasi (populasi dibagi dengan ukuran sampel yang diperlukan (n) dan sampel diperoleh dengan cara mengambil setiap subjek ke-n).

Contoh, populasi 100, ukuran sampel 10. Ukuran sampel, 100/10 = 10. Selanjutnya, pilih nomor antara 1 dan 10, misalnya 5. Kemudian pilih yang ke 10, setelah itu hingga 10 dipilih 5, 15, 25, 35, 45, 55, 65, 75, 85, 95.

b. **Sampling Wilayah**

Sampling klaster dalam suatu wilayah.

Contoh, sebuah stasiun radio melakukan survei profil dan perilaku pendengar radio. Gunakan peta kota, lalu kecamatannya, kelurahan, RW dan RT yang terpilih. Selanjutnya sampel dipilih secara acak dari setiap klaster tersebut.

c. **Sampling Kemudahan**

Untuk mendapatkan informasi dengan cepat, mudah dan murah. Prosedurnya: langsung menghubungi unit-unit sampling yang mudah dijumpai, seperti mahasiswa di suatu kelas, jemaah tempat-tempat ibadah, rekan-rekan, para tetangga, dll. Sering kali teknik sampling ini dilakukan untuk menguji kuesioner atau digunakan dalam penelitian eksplorasi.

d. **Sampling Pertimbangan**

Didasarkan pada kriteria-kriteria tertentu.

Misalnya dalam suatu penelitian tentang masalah sumber daya manusia, peneliti mungkin hanya ingin memperoleh informasi dari pegawai-pegawai yang memiliki karakteristik tertentu. Dalam kaitannya dengan sampling pertimbangan dikenal juga sampling ahli (*expert sampling*) dan sampling bertujuan (*purposive sampling*). Kendala yang dihadapi dalam penggunaan sampling pertimbangan ini adalah tuntutan adanya kejelian dari peneliti dalam mendefinisikan populasi dan membuat pertimbangannya. Pertimbangan atau *judgement* harus masuk akal dan relevan dengan maksud penelitian.

e. **Sampling Kuota**

Bentuk lain sampling pertimbangan, karakteristik-karakteristik tertentu yang relevan yang menjelaskan dimensi-dimensi populasi. Dalam hal ini, distribusi populasi harus diketahui. Misal, sampel sebanyak 1000 orang penduduk kota Bandung. Jika diketahui penyebaran penduduk secara geografis, sampelnya dapat ditarik persentase distribusi yang sama.a, bahkan pada kondisi tertentu, hasil penelitian dapat menyamai hasil penelitian yang dilakukan dengan teknik sampling probabilitas.

Populasi terhingga dan tak tehingga

Populasi terhingga (finite population) adalah populasi yang jumlah seluruh anggotanya tetap dan dapat didaftar. Populasi tak terhingga (infinite population) memiliki anggota yang banyaknya tak terhingga

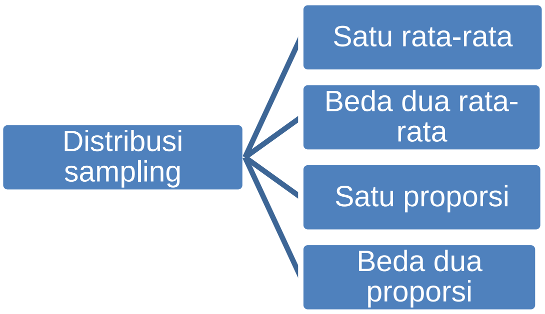
Contoh.

Jika kita memeriksa rata-rata harian banyaknya produk cacat di sebuah pabrik selama 12 bulan terakhir, maka populasi yang diperoleh adalah populasi terhingga yang meliputi produk cacat dari semua jalur produksi di pabrik itu.

Jika kita mengukur kecepatan prosesor komputer yang dibuat oleh sebuah perusahaan tertentu maka populasi yang diperoleh adalah populasi tak terhingga, karena produk tersebut akan terus diproduksi dan dikembangkan di masa-masa mendatang

DISTRIBUSI SAMPLING

Pada modul ini yang dibahas mengenai distribusi sampling digambarkan dalam skema berikut.



Distribusi Sampling Rata-Rata

Jika populasinya A, B, C, kemudian diambil 2 sampel, maka diperoleh 3 kombinasi sampel, yaitu AB, AC dan BC. Jika populasi =60, ambil 10 sampel maka kita mempunyai kombinasi 60 dari 10.

Ukuran populasi (N) biasanya sangat besar dan ukuran sampel (n) relatif lebih kecil. Jika n sampel diambil dari N populasi maka memiliki kombinasi N terhadap n $(C(N,n))$. Setiap kombinasi sampel memiliki ukuran (statistik sampel), misalnya rata-rata sampel. Maka jika kita mengambil n sampel dari N populasi, kita akan memiliki rata-rata sampel yang cukup banyak.

Distribusi sampling rata-rata merupakan distribusi normal, yang berbentuk lonceng, simetris dan memiliki rata-rata dan deviasi standar.

Contoh.

PT Green Bay Packer memiliki 7 karyawan bagian produksi (dianggap sebagai populasi) dengan keterangan upah per jam setiap karyawan seperti berikut:

Karyawan	Upah/Jam
Joe	7
Sam	9
Sue	8
Bob	8
Jan	7
Art	8
Teed	9

Penyelesaian:

Seorang peneliti ingin mengetahui rata-rata upah per jam karyawan di perusahaan tersebut. Untuk melakukannya, ia dapat menggunakan dua cara, yaitu :

a. Meneliti seluruh populasi

   Rata-rata populasinya adalah:

$$\mu = \frac{7+9+8+8+7+8+9}{7} = 8$$

b. Meneliti sampel

   Misal, ambil 4 karyawan, maka ada 35 kombinasi sampel yang mungkin terambil, yaitu dari perhitungan

$$C(7,4) = \frac{7!}{(7-4)!\,4!} = \frac{7!}{3!\,4!} = \frac{7\times 6\times 5\times 4!}{3\times 2\times 1\times 4!} = 35$$

   Masing-masing kombinasi sampel memiliki statistik sampel atau rata-rata sampel. 35 sampel tersebut disajikan pada tabel berikut ini

Rata-rata sampel	Frekuensi	Probabilitas
7.5	3	3/35 = 0.0857
7.75	8	8/35 = 0.2286
8	13	13/35 = 0.3714
8.25	8	8/35 = 0.2286
8.5	3	3/35 = 0.0837
	35	1

Distribusi sampling tersebut memiliki rata-rata

$$\mu_x = \sum \frac{x_i}{n}$$

dan standar deviasi

$$\sigma_x = \sqrt{\sum \frac{(x_i - \mu_x)^2}{n-1}}$$

Deviasi standar distribusi sampling rata-rata disebut juga galat baku mean (*standard error of mean*).

Dalil Batas Memusat (*The Central Limit Theorem*)

Dalil yang menyatakan bahwa untuk suatu populasi dengan rata-rata μ dan varian σ: distribusi sampling rata-rata dari semua kemungkinan sampel berukuran n yang diambil dari populasi akan terdistribusi secara normal dengan rata-rata μ_x sama dengan rata-rata populasi (μ) dan deviasi standar $\sigma_{x,}$ dengan deviasi standar populasi dibagi akar n atau $\frac{\sigma}{\sqrt{n}}$, dengan asumsi bahwa ukuran sampel cukup besar.

Dengan kata lain, dalam pemilihan sampel acak sederhana dengan ukuran n dari suatu populasi yang berasal dari distribusi apapun (binomial, poisson, dsb), distribusi dari rata-rata sampel dapat didekati dengan distribusi probabilitas normal untuk ukuran sampel yang besar. Beberapa hal penting yang perlu diingat dari dalil tersebut adalah sebagai berikut.

Jika ukuran sampel (n) cukup besar, distribusi rata-rata sampel akan mendekati normal, tidak peduli apakah populasinya terdistribusi secara normal atau tidak, dengan

$$\mu_x = \mu$$

dan

$$\sigma_x = \frac{\sigma}{\sqrt{n}}$$

keterangan :

μ_x = rata-rata dari distribusi sampling rata-rata
μ = rata-rata populasi
σ_x = deviasi standar dari distribusi sampling rata-rata
σ = deviasi standar populasi

$n =$ ukuran sampel

Tidak ada angka yang pasti tentang “ukuran sampel yang cukup besar”, tetapi biasanya angka $n > 30$ dianggap cukup besar.

Contoh.

Bank Pasti Aman menghitung tabungan seluruh nasabahnya. Setelah penghitungan, bank tersebut mendapati bahwa rata-rata tabungan setiap nasabahnya sebesar Rp2.000, dengan deviasi standar Rp600, apabila seorang peneliti mengambil sampel sebanyak 100 nasabah, berapa probabilitas jika :

a. Rata-rata sampel akan lebih kecil dari Rp1.900
b. Rata-rata sampel akan lebih kecil dari Rp2.050
c. Rata-rata sampel akan terletak antara Rp1.900 dan Rp2.050

Penyelesaian.

Dari soal diketahui bahwa

$n = 100$
$\mu = 2000$
$\sigma = 500$
Sehingga diperoleh

$$\mu_x = \mu = 2000$$

dan

$$\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{600}{\sqrt{100}} = 60$$

Dengan menggunakan perhitungan distribusi normal, yaitu dengan rumus

$$Z = \frac{x - \mu_x}{\sigma_x}$$

diperoleh sebagai berikut.

a. Kasus dimana rata-rata sampel lebih kecil dari Rp1.900

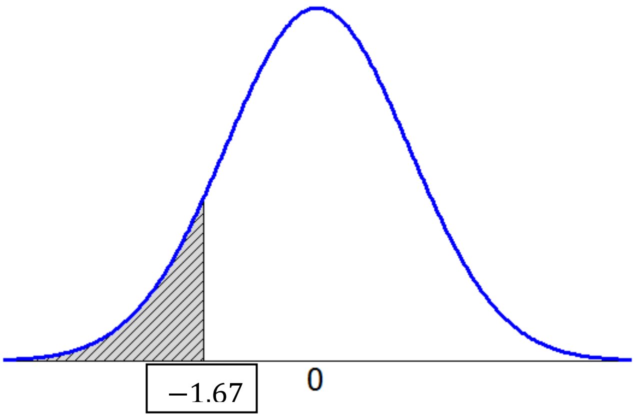
$$P(x < 1900) = P\left(Z < \frac{x - \mu_x}{\sigma_x}\right)$$

$$= P\left(Z < \frac{1900 - 2000}{60}\right)$$

$$= P(Z < -1{,}67)$$

$$= 0{,}0475$$

Cara perhitungan dan ilustrasi untuk $P(Z < -1{,}67)$ adalah sebagai berikut.



$P(Z < -1{,}67)$ adalah luas daerah yang kurang dari $-1{,}67$. Lihat tabel binomial pada Lampiran 1, silahkan dicari nilai dari -1,6 pada baris keberapa dan 0,07 pada kolom keberapa. Sehingga diperoleh nilai $P(Z < -1{,}67) = 0{,}0475$.

Jadi, probabilitas rata-rata sampel lebih kecil dari Rp1.900 adalah 4,75%.

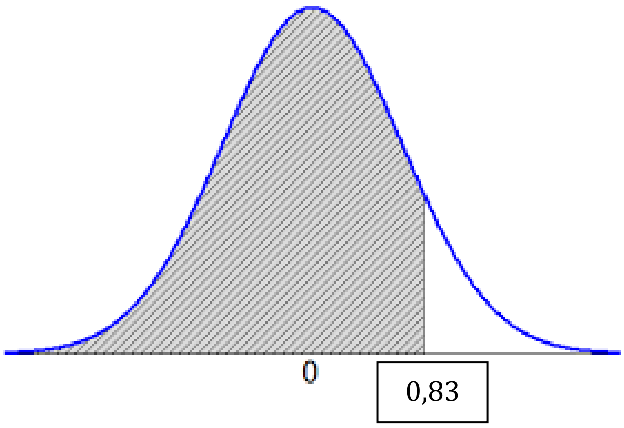
b. Kasus dimana rata-rata sampel lebih kecil dari Rp2.050

$$P(x < 2050) = P\left(Z < \frac{x - \mu_x}{\sigma_x}\right)$$

$$= P\left(Z < \frac{2050 - 2000}{60}\right)$$

$$= P(Z < 0{,}83)$$

$$= 0{,}7967$$



Perhitungan $P(Z < 0{,}83)$, identik dengan perhitungan $P(Z < -1{,}67)$.

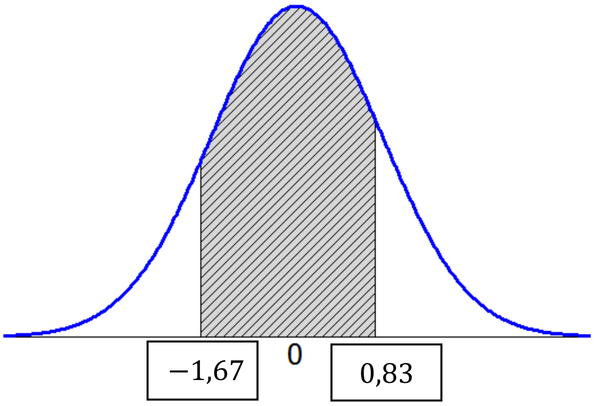
Jadi, probabilitas rata-rata sampel lebih kecil dari Rp2.050 adalah 79,67%.

c. Kasus dimana rata-rata sampel terletak antara Rp1.900 dan Rp2.050

$$P(1900 < x < 2050) = P\left(\frac{x - \mu_x}{\sigma_x} < Z < \frac{x - \mu_x}{\sigma_x}\right)$$

$$= P\left(\frac{1900 - 2000}{60} < Z < \frac{2050 - 2000}{60}\right)$$

$$= P(-1{,}67 < Z < 0{,}83)$$



Perhitungan nilai dari $P(-1{,}67 < Z < 0{,}83)$ adalah sebagai berikut.

Dari kurva diatas diketahui bahwa luasan yang berada diantara -1,67 dan 0,83 akan diperoleh jika dihitung luasan yang kurang dari 0,83 dan luasan yang kurang dari -1,67. Sehingga diperoleh

$$
\begin{aligned}
P(-1{,}67 < Z < 0{,}83) &= P(Z < 0{,}83) - P(Z < -1{,}67) \\
&= 0{,}7967 - 0{,}0475 \\
&= 0{,}7967 - 0{,}0475 \\
&= 0{,}7492
\end{aligned}
$$

Jadi, probabilitas rata-rata sampel terletak antara Rp1.900 dan Rp2.050 adalah 74,92%.

Distribusi Sampling Beda Rata-rata

Misalkan populasi 1 adalah N_1 dan sampel 1 adalah n_1, maka diperoleh $C(N_1, n_1)$. Misalkan Populasi 2 adalah N_2 dan sampel 2 adalah n_2 maka diperoleh $C(N_2, n_2)$. Dengan kata lain, kita memiliki rata-rata sampel dari populasi 1 dan rata-rata sampel dari populasi 2 yang cukup banyak.

Jika $x = x_1 - x_2$, kita akan memiliki X yang banyak sekali yang membentuk suatu distribusi normal yang disebut distribusi sampling beda rata-rata dengan rata-rata $\mu_{x_1-x_2}$ dan deviasi standar atau $\sigma_{x_1-x_2}$. Jika kita mengurangi x_1 dengan x_2, kita akan mendapat variabel $x_1 - x_2$ yang banyak sekali yang membentuk **distribusi normal**.

Menurut dalil batas memusat, diperoleh:

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

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

Contoh :

Lampu pijar merek ampuh memiliki rata-rata daya tahan 4500 jam dengan deviasi standar 500 jam, sedangkan lampu pijar merek baik memiliki rata-rata daya tahan 4000 jam dengan deviasi standar 400 jam. Jika diambil sampel masing-masing 100 buah lampu pijar dan diteliti, berapa probabilitas bahwa selisih rata-rata daya tahan kedua lampu pijar tersebut lebih besar dari 600 jam?

Penyelesaian.

Dari soal diketahui bahwa

$\mu_1 = 4500$, $\sigma_1 = 500$

$\mu_2 = 4000$, $\sigma_2 = 400$

$n_1 = n_2 = 100$

dan ditanya: $P(x_1 - x_2 > 600)$?

Jawab:

Pertama, dilakukan perhitungan $\mu_{x_1-x_2}$ dan $\sigma_{x_1-x_2}$, sehingga diperoleh

$$\mu_{x_1-x_2} = \mu_1 - \mu_2 = 4500 - 4000 = 500$$

dan

$$\begin{aligned}\sigma_{x_1-x_2} &= \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma_2^2}{n_2}}\\ &= \sqrt{\frac{500^2}{100} + \frac{400^2}{100}}\\ &= \sqrt{4100}\\ &= 64{,}031\end{aligned}$$

Selanjutnya dihitung nilai dari $P(x_1 - x_2 > 600)$ dengan menggunakan distribusi normal standar, yaitu dengan rumus

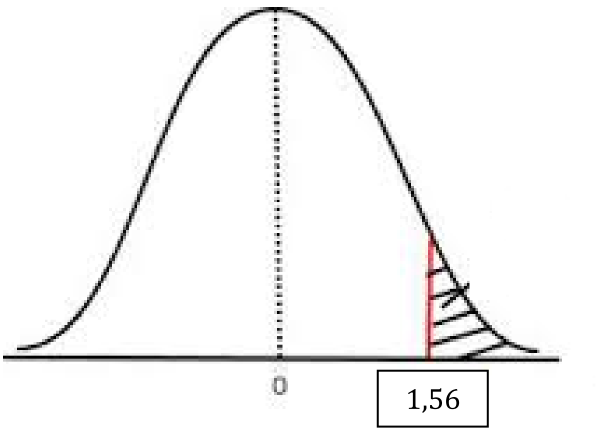
$$Z = \frac{(x_1 - x_2) - \mu_{x_1-x_2}}{\sigma_{x_1-x_2}}$$

Sehingga diperoleh

$$P(x_1 - x_2 > 600) = P\left(Z > \frac{(x_1 - x_2) - \mu_{x_1-x_2}}{\sigma_{x_1-x_2}}\right)$$

$$= P\left(Z > \frac{600 - 500}{64{,}031}\right)$$

$$= P(Z > 1{,}56)$$



Kita harus mengingat kembali, bahwa luasan seluruh kurva nomal adalah 1. Jika kita ingin menghitung luasan dari daerah yang lebih dari 1,56, maka kita cukup mengurangi semua luasan kurva nurmal dengan luasan dari daerah yang kurang dari 1,56. Sehingga diperoleh

$$\begin{aligned} P(Z > 1{,}56) &= 1 - P(Z < 1{,}56) \\ &= 1 - 0{,}9406 \\ &= 0{,}0594 \end{aligned}$$

Jadi, probabilitas bahwa selisih rata-rata daya tahan kedua lampu pijar tersebut lebih besar dari 600 jam adalah 5,94%.

Distribusi Sampling Proporsi

Misalkan proporsi populasi dinotasikan dengan p dengan $p = \frac{x}{N}$.
dimana
$x =$ jumlah item proporsi
$N =$ adalah jumlah seluruh item.

Sebagai contoh, total mahasiswa adalah 100 orang, jika 30 mahasiswa diantaranya merokok, proporsi mahasiswa yang merokok adalah 30/100 atau 30%.

Misalkan populasi adalah N dan sampel adalah n, berarti kita telah memiliki p (proporsi sampel) sebanyak $C(N,n)$. P yang dimaksud berjumlah sangat besar dan membentuk distribusi normal dengan rata-rata μ_p dan deviasi standar σ_p. Dimana

$$\mu_p = p = \frac{x}{N}$$

Dan

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}}$$

Contoh :

Dari 1000 mobil yang yang diproduksi, diketahui 100 diantaranya cacat. Jika diambil sampel acak sebanyak 500 buah mobill dari populasi tersebut dan diteliti, berapa probabilitas besar proporsi mobil yang cacat lebih besar dari 12%?

Penyelesaian.

Diketahui

$N = 1000$
$n = 500$
x = banyaknya mobil yang rusak $= 100$
Ditanya $P(p > 0{,}12)$?

Jawab:

pertama-tama, dihitung nilai dari μ_p dan σ_p. diperoleh

$$\mu_p = p = \frac{x}{N} = 0{,}1$$

dan

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0{,}1(1-0{,}1)}{500}} = 0{,}013$$

Selanjutnya dihitung nilai dari $P(p > 0{,}12)$ dengan menggunakan distribusi normal standar, yaitu dengan rumus

$$Z = \frac{p - \mu_p}{\sigma_p}$$

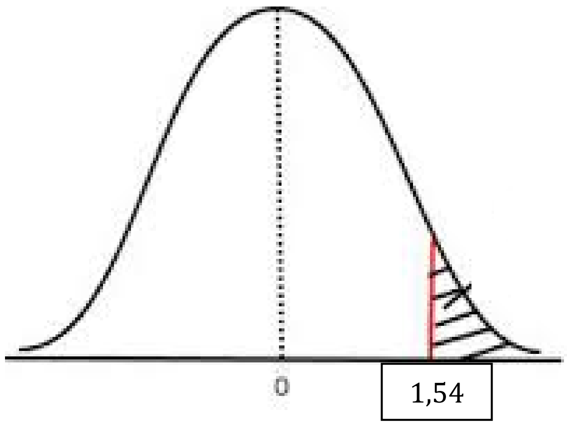
Sehingga diperoleh

$$P(p > 0{,}12) = P\left(Z > \frac{p - \mu_p}{\sigma_p}\right)$$

$$= P\left(Z > \frac{0{,}12 - 0{,}1}{0{,}0134}\right)$$

$$= P(Z > 1{,}49)$$

$$= P(Z > 1{,}54)$$



$$P(Z > 1{,}56) = 1 - P(Z < 1{,}54) = 1 - 0{,}9382 = 0{,}0618$$

Jadi, probabilitas besar proporsi mobil yang cacat lebih besar dari 12% adalah 6,18%.

Distribusi Sampling Beda Proporsi

Misalkan populasi 1 adalah N_1 dan sampel 1 adalah n_1 maka terdapat $C(N_1, n_1)$ dengan proporsi p_1. Sedangkan populasi 2 adalah N_2 dan sampel 2 adalah n_2 maka terdapat $C(N_2, n_2)$ dengan proporsi p_2. Selisih dari p_1 dan p_2 membentuk distribusi normal dengan rata-ratanya adalah

$$\mu_{p_1-p_2} = p_1 - p_2$$

dan deviasi standarnya adalah

$$\sigma_{p_1-p_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

atau

$$\sigma_{p_1-p_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \times \sqrt{\frac{N-n}{N-1}}$$

untuk populasi terbatas.

Contoh.

Berdasarkan sebuah penelitian dengan sampel masing-masing 100 orang dari populasi perokok dan bukan perokok, diketahui bahwa 1 orang yang tidak merokok terkena TBC dan 5 orang yang merokok terkena TBC. Berapa probabilitas bahwa selisih proporsi populasi perokok dan populasi bukan perokok yang terkena TBC lebih besar dari 5%?

Penyelesaian.

Diketahui

$n_1 = 100$
$n_2 = 100$
x_1 = banyaknya sampel dari perokok yang terkena TBC $= 5$
x_2 = banyaknya sampel yang bukan perokok yang terkena TBC $= 1$
p_1 = proporsi populasi perokok yang terkena TBC $= \frac{5}{100} = 0{,}05$
p_2 = proporsi populasi bukan perokok yang terkena TBC $= \frac{1}{100} = 0{,}01$
yang ditanya : $P(p_1 - p_2 > 0{,}05)$?

Jawab.

Pertama-tama, dilakukan perhitungan rata-rata dan deviasi standarnya.

$$\mu_{p_1-p_2} = p_1 - p_2 = 0{,}05 - 0{,}01 = 0{,}04$$

$$\begin{aligned}\sigma_{p_1-p_2} &= \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \\ &= \sqrt{\frac{0{,}05(1-0{,}05)}{100} + \frac{0{,}01(1-0{,}01)}{100}} \\ &= 0{,}024\end{aligned}$$

Selanjutnya dihitung nilai dari $P(p_1 - p_2 > 0{,}05)$ dengan menggunakan distribusi normal standar, yaitu dengan rumus

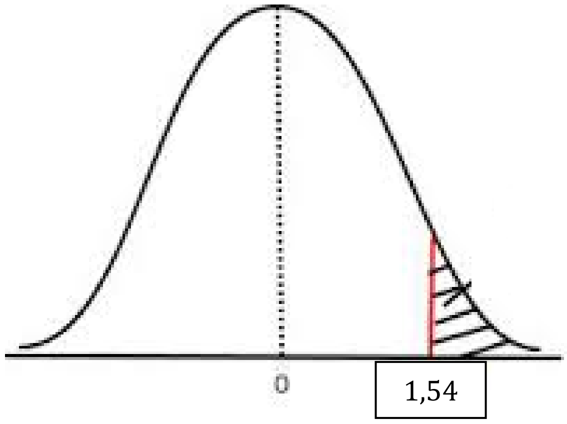
$$Z = \frac{(p_1 - p_2) - \mu_{p_1 - p_2}}{\sigma_{p_1 - p_2}}$$

Sehingga diperoleh

$$P(p_1 - p_2 > 0{,}05) = P\left(Z > \frac{(p_1 - p_2) - \mu_{p_1 - p_2}}{\sigma_{p_1 - p_2}}\right)$$

$$= P\left(Z > \frac{0{,}05 - 0{,}04}{0{,}024}\right)$$

$$= P(Z > 0{,}42)$$



$$P(Z > 0{,}42) = 1 - P(Z < 0{,}42)$$
$$= 1 - 0{,}6628$$
$$= 0{,}3372$$

probabilitas bahwa selisih proporsi populasi perokok dan populasi bukan perokok yang terkena TBC lebih besar dari 5% adalah 33,72%.

Kesimpulan:

1. Sampling dilakukan untuk jumlah populasi yang sangat besar.
2. Sampling Objektif dilakukan jika semua elemen populasi memiliki peluang yang sama sebagai sampel, sementara sampling subjektif dilakukan untuk elemen yang tidak memiliki peluang yang sama sebagai sampel.
3. Karakteristik sampel sebagai hasil sampling bisa digunakan untuk menyimpulkan populasi.

Referensi:

Douglas C. Montgomery, George C. Runger, 2003, *Applied Statistic and Probability for Engineer*, third edition, John Wiley and Son Inc.

http://staff.ui.ac.id/system/files/users/harinaldi.d/material/statistik-05.pdf

J. Supranto, M.A. ,2001, *Statistika Teori dan Aplikasi*, Erlangga, Jakarta.

Johson & Wichern, 2007, *Applied multivariate statistical analysis*, Upper Saddle River: Pearson Prentice Hall.

Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers and Keying Ye, 2007, *Probabilitiy and Statistics for Engineers and Scientists,* 8th edition, Pearson Prentice Hall.

Sharma, Subhash, 1996, *Applied Multivariate Techniques*, John Willey & Son, Inc., USA.

Singgih Santoso, 2014, *Panduan Lengkap SPSSversi 20*, Alex Media Komputindo.

Lampiran 1.

Tabel Normal.

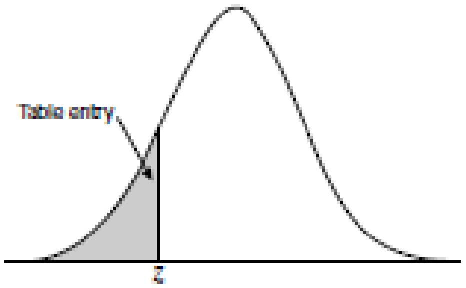


Table entry for *z* is the area under the standard normal curve to the left of *z*.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
–3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
–3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
–3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
–3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
–3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
–2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
–2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
–2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
–2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
–2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
–2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
–2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
–2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
–2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
–2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
–1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
–1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
–1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
–1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
–1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
–1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
–1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
–1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
–1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
–1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
–0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
–0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
–0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
–0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
–0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
–0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
–0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
–0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
–0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
–0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

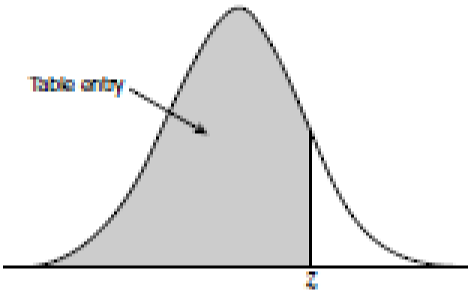


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998