

Metode Sampling dan Teorema Central Limit

Tjipto Juwono, Ph.D.

November 2017



SURYA
UNIVERSITY

Mengapa Perlu Sampling?

Contoh

Kita ingin mengetahui elektabilitas para calon presiden Indonesia. Bagaimana caranya?

- 1 Mewawancarai seluruh rakyat Indonesia? Terlalu lama, dan terlalu mahal.
- 2 Lagi pula: hasil dari sampling sudah cukup memadai untuk memperoleh informasi tentang elektabilitas para calon presiden itu.

Jadi, bukan saja terlalu lama dan terlalu mahal, tetapi juga: *tidak perlu*.

Mengapa Perlu Sampling?

Contoh

Kita ingin mengetahui kualitas satu truk jeruk. Apakah bagus dan manis?

- 1 Kita tidak dapat mencicipi seluruh jeruk. Proses mencicipi adalah proses yang *destruktif*.
- 2 Sekali lagi: hasil dari sampling sudah cukup memadai untuk memperoleh informasi tentang jeruk satu truk itu.

Jadi, bukan saja bersifat destruktif pada populasi, tetapi juga: *tidak perlu*.

Mengapa Perlu Sampling?

Contoh

Kita ingin mengetahui kualitas air di sebuah danau yang besar (misalnya danau Toba).

- 1 Kita tidak dapat menguji seluruh air di danau itu.
- 2 Sekali lagi: hasil dari sampling sudah cukup memadai untuk memperoleh informasi tentang kualitas air di danau tadi.

Jadi, bukan saja tidak mungkin untuk memeriksa seluruh populasi, tetapi juga: *tidak perlu*.

Mengapa Perlu Sampling?

- 1 Menguji atau mengkontak seluruh populasi **membutuhkan waktu yang terlalu lama**.
- 2 Menguji atau mengkontak seluruh populasi **membutuhkan biaya yang terlalu mahal**.
- 3 Menguji atau mengkontak seluruh populasi **mustahil secara fisis**.
- 4 Beberapa jenis pengujian bersifat **destruktif**.
- 5 Hasil dari sampling sudah cukup **memadai** untuk merepresentasikan seluruh populasi.

Probability Sampling

Probability Sampling

Probability Sample: adalah sample yang dipilih sedemikian sehingga kita mengetahui berapa kemungkinan setiap item atau orang di dalam populasi untuk dimasukkan ke dalam sample.

Beberapa metode probability sampling

Metode Probability Sampling

- Simple Random Sample
- Systematic Random Sampling
- Stratified Random Sampling
- Cluster Sampling

Simple Random Sample

Simple Random Sample

Suatu sample yang dipilih sedemikian rupa sehingga setiap anggota populasi mempunyai kemungkinan yang sama untuk dipilih sebagai sample.

Simple Random Sample

Contoh

Suatu populasi terdiri dari 845 karyawan dari suatu perusahaan. Suatu sample yang terdiri dari 52 karyawan akan dipilih dari populasi itu. Bagaimana caranya?

Simple Random Sample

Contoh

Suatu populasi terdiri dari 845 karyawan dari suatu perusahaan. Suatu sample yang terdiri dari 52 karyawan akan dipilih dari populasi itu. Bagaimana caranya?

Nama-nama dari setiap karyawan ditulis pada selembar kertas, yang lalu dimasukkan ke dalam kotak. Setelah seluruhnya diacak, pilihan pertama dilakukan dengan mengambil selembar kertas dari kotak tanpa melihatnya. Proses ini diulang sampai diperoleh sample terdiri dari 52 karyawan.

Simple Random Sample

Contoh

Suatu populasi terdiri dari 845 karyawan dari suatu perusahaan. Suatu sample yang terdiri dari 52 karyawan akan dipilih dari populasi itu. Bagaimana caranya?

Setiap karyawan diwakili oleh satu nomor 1 s/d 845, lalu gunakan tabel bilangan random untuk memilih secara acak.

Simple Random Sample

Tabel Bilangan Random

50525	57454	28455	68226	34656	38884	39018
72507	53380	53827	42486	54465	71819	91199
34986	74297	00144	38676	89967	98869	39744
68851	27305	03759	44723	96108	78489	18910
06738	62879	03910	17350	49169	03850	18910
11448	10734	05837	24397	10420	16712	94496
		↓	↓	↓	↓	↓
		Starting point	Second employee	Third employee	Fourth employee	

Systematic Random Sample

Systematic Random Sample

Anggota-anggota diatur menurut suatu urutan. Titik awal dipilih secara random kemudian setiap anggota ke- k dipilih untuk sample.

Systematic Random Sample

Contoh

Suatu populasi terdiri dari 845 karyawan dari suatu perusahaan. Suatu sample yang terdiri dari 52 karyawan akan dipilih dari populasi itu. Gunakan systematic random sampling untuk melakukannya!

Systematic Random Sample

Contoh

Suatu populasi terdiri dari 845 karyawan dari suatu perusahaan. Suatu sample yang terdiri dari 52 karyawan akan dipilih dari populasi itu. Gunakan systematic random sampling untuk melakukannya!

- Hitung k :

$$k = \frac{\text{Ukuran populasi}}{\text{Ukuran sample}} = \frac{845}{52} \approx 16 \quad (1)$$

- Jika k bukan bilangan bulat, bulatkan ke bawah.
- Pilih nama pertama secara random, lalu pilih setiap nama ke-16 dalam daftar.

Stratified Random Sample

Stratified Random Sample

Populasi dibagi-bagi atas sejumlah golongan/kelompok yang disebut *strata*. (Jamak: *strata*, tunggal: *stratum*). Kemudian sample diambil dari setiap *stratum*. Metode ini berguna jika suatu populasi memang jelas-jelas dapat dibagi-bagi atas sejumlah golongan berdasarkan karakteristik tertentu.

Stratified Random Sample

Contoh

Misalkan kita ingin mempelajari pengeluaran untuk iklan dari 352 perusahaan untuk menentukan apakah perusahaan dengan ROE yang tinggi (salah satu ukuran untuk profitabilitas) membelanjakan lebih banyak untuk iklan daripada perusahaan-perusahaan dengan ROE rendah, atau defisit.

Stratified Random Sample

Stratum	Profitability (return on equity)	Number of Firms	Relative Frequency	Number Sampled
1	30 percent and over	8	0.02	1*
2	20 up to 30 percent	35	0.10	5*
3	10 up to 20 percent	189	0.54	27
4	0 up to 10 percent	115	0.33	16
5	Deficit	5	0.01	1
Total		352	1.00	50

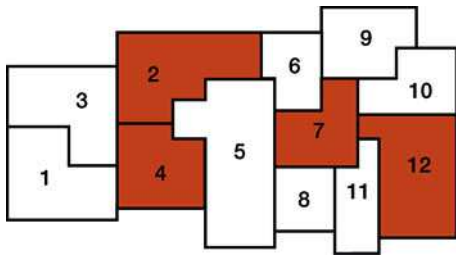
Cluster Sampling

Cluster Sampling

Populasi dibagi-bagi atas cluster-cluster dengan menggunakan batas geografis alamiah, atau batas-batas lainnya. Kemudian sejumlah cluster dipilih secara random, selanjutnya sample dipilih dengan memilih secara random dari setiap cluster

Cluster Sampling

Misalkan kita ingin mempelajari sikap rakyat dari suatu propinsi tentang kebijakan presiden. Bagilah propinsi itu atas cluster-cluster (misalkan berdasarkan kabupaten) lalu pilih beberapa cluster secara random (dalam contoh di samping 4 cluster dipilih secara random). Lalu ambil sampel dari setiap cluster yang dipilih dengan memilih penduduknya secara random.



Sampling Error

Sampling Error

Adalah selisih antara statistik sample dan parameter populasi yang bersesuaian.

$$\begin{aligned}\bar{X} &- \mu \\ S &- \sigma \\ S^2 &- \sigma^2 \\ p &- \pi\end{aligned}\tag{2}$$

Distribusi Sampling dari Sampling Mean

Distribusi Sampling dari Sampling Mean

Distribusi probabilitas yang terdiri dari semua sample mean yang mungkin dari suatu ukuran sample tertentu yang diambil dari suatu populasi.

Distribusi Sampling dari Sampling Mean

Contoh: Sebuah pabrik kecil mempunyai 7 karyawan (anggap sebagai populasi). Upah per jam diberikan pada tabel berikut.

Employee	Hourly Earnings	Employee	Hourly Earnings
Joe	\$7	Jan	\$7
Sam	7	Art	8
Sue	8	Ted	9
Bob	8		

Pertanyaan:

- 1 Hitung population mean.
- 2 Carilah distribusi sampling dari sample mean untuk ukuran sample $n = 2$.
- 3 Hitung mean dari distribusi sampling!
- 4 Hal-hal apa saja yang dapat diamati mengenai populasi dan sampling?

Distribusi Sampling dari Sampling Mean

- ① **Hitung population mean.** Population mean:

$$\begin{aligned}\mu &= \frac{\sum X}{N} \\ &= \frac{\$7 + \$7 + \$8 + \$7 + \$8 + \$9}{7} \\ &= \$7.71\end{aligned}\tag{3}$$

Distribusi Sampling dari Sampling Mean

- ② **Carilah distribusi sampling dari sample mean untuk ukuran sample $n=2$.** Untuk memperoleh distribusi sampling dari sample mean, kita perlu memilih semua sample-sample berukuran $n = 2$ yang mungkin dari populasi. Jumlah sample-sample berukuran $n = 2$ yang mungkin dihitung dengan kombinasi

$$\begin{aligned} {}_N C_n &= \frac{N!}{n!(N-n)!} \\ &= \frac{7!}{2!(7-2)!} \\ &= 21 \end{aligned} \tag{4}$$

Distribusi Sampling dari Sampling Mean

Sample	Employees	Hourly Earnings	Sum	Mean	Sample	Employees	Hourly Earnings	Sum	Mean
1	Joe, Sam	\$7, \$7	\$14	\$7.00	12	Sue, Bob	\$8, \$8	\$16	\$8.00
2	Joe, Sue	7, 8	15	7.50	13	Sue, Jan	8, 7	15	7.50
3	Joe, Bob	7, 8	15	7.50	14	Sue, Art	8, 8	16	8.00
4	Joe, Jan	7, 7	14	7.00	15	Sue, Ted	8, 9	17	8.50
5	Joe, Art	7, 8	15	7.50	16	Bob, Jan	8, 7	15	7.50
6	Joe, Ted	7, 9	16	8.00	17	Bob, Art	8, 8	16	8.00
7	Sam, Sue	7, 8	15	7.50	18	Bob, Ted	8, 9	17	8.50
8	Sam, Bob	7, 8	15	7.50	19	Jan, Art	7, 8	15	7.50
9	Sam, Jan	7, 7	14	7.00	20	Jan, Ted	7, 9	16	8.00
10	Sam, Art	7, 8	15	7.50	21	Art, Ted	8, 9	17	8.50
11	Sam, Ted	7, 9	16	8.00					

Distribusi Sampling dari Sampling Mean

Distribusi Sampling dari Sampling Mean (dalam bentuk tabel)

Sample Mean	Number of Means	Probability
\$7.00	3	.1429
7.50	9	.4285
8.00	6	.2857
8.50	3	.1429
	<u>21</u>	<u>1.0000</u>

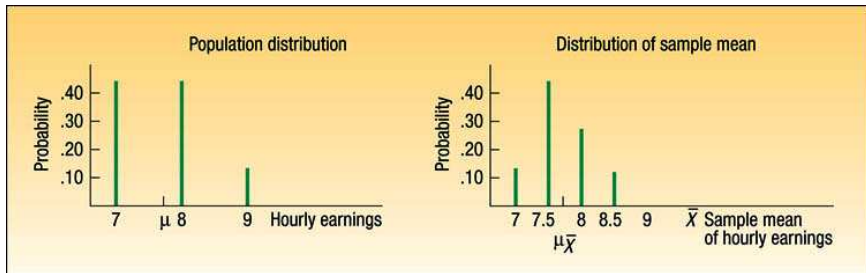
Distribusi Sampling dari Sampling Mean

3 Hitung mean dari distribusi sampling.

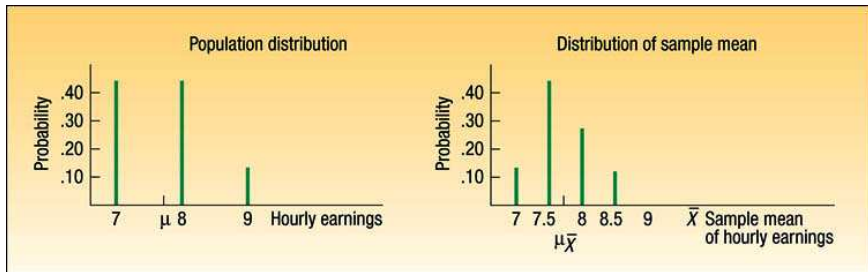
$$\begin{aligned}\mu_{\bar{X}} &= \frac{\text{Jumlah dari semua mean sample}}{\text{Jumlah sample}} \\ &= \frac{\$7.00 + \$7.50 + \dots + \$8.50}{21} \\ &= \$7.71\end{aligned}\tag{5}$$

Distribusi Sampling dari Sampling Mean

- 4 Hal-hal apa saja yang dapat diamati mengenai populasi dan sampling?



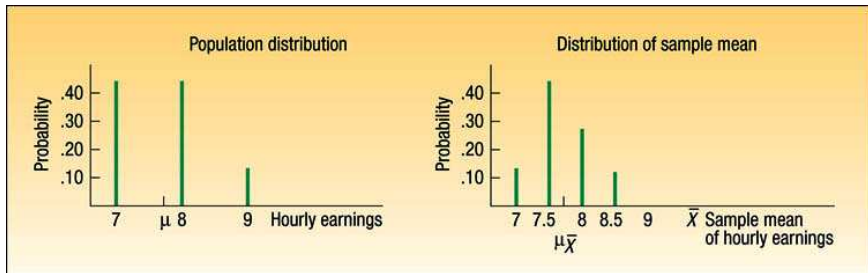
Distribusi Sampling dari Sampling Mean



Mean dari populasi = Mean dari distribusi sample mean

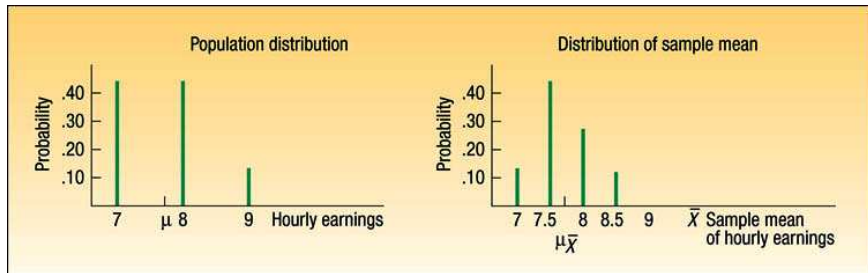
$$\mu = \mu_{\bar{X}} = \$7.71$$

Distribusi Sampling dari Sampling Mean



Sebaran dari distribusi sample mean lebih kecil dari sebaran populasi. Jika ukuran sample ditambah, sebaran dari distribusi sample mean akan bertambah kecil.

Distribusi Sampling dari Sampling Mean



Bentuk sebaran dari distribusi sample mean berbeda dari bentuk sebaran populasi. Sebaran dari distribusi sample mean cenderung mendekati bentuk distribusi normal.

Central Limit Theorem

Central Limit Theorem

Central Limit Theorem

Jika semua sample dengan ukuran tertentu diambil dari suatu populasi, maka distribusi sampling dari sample mean akan mendekati distribusi normal. Aproksimasi ini akan menjadi lebih baik dengan sample-sample yang berukuran lebih besar.

Central Limit Theorem

- 1 Jika populasinya mengikuti distribusi normal, maka untuk berapapun ukuran sampelnya, distribusi sampling dari sample mean akan selalu berupa distribusi normal.

Central Limit Theorem

- 1 Jika populasinya mengikuti distribusi normal, maka untuk berapapun ukuran sampelnya, distribusi sampling dari sample mean akan selalu berupa distribusi normal.

Central Limit Theorem

- ② Jika distribusi populasi simetris (tapi tidak normal), maka bentuk distribusi normal pada distribusi sampling dari sample mean akan mulai nampak pada ukuran sample $n \geq 10$.

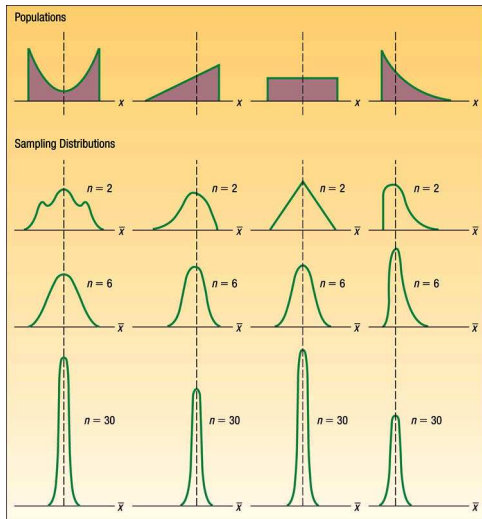
Central Limit Theorem

- 3 Jika distribusi populasi tidak simetris, maka bentuk distribusi normal pada distribusi sampling dari sample mean akan mulai nampak pada ukuran sample $n \geq 30$.

Central Limit Theorem

- ④ Mean dari distribusi sampling dari sample mean adalah μ .
- ⑤ Variance dari distribusi sampling dari sample mean adalah μ^2/n .

Central Limit Theorem



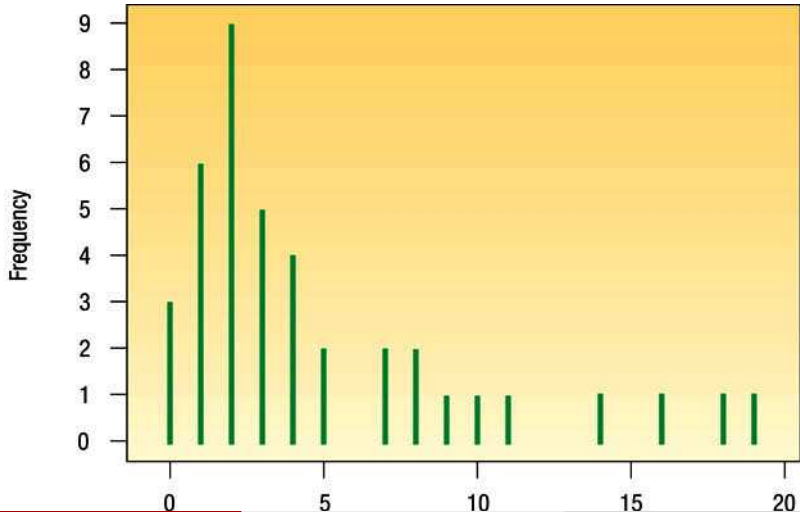
Central Limit Theorem

Contoh: Berikut ini adalah lama kerja dari 40 karyawan di sebuah perusahaan. Jika suatu sample diambil dari populasi ini, apa ekspektasi dari mean sample itu? Bagaimana perbandingan antara distribusi sampling dari sample mean dengan distribusi populasinya?

11	4	18	2	1	2	0	2	2	4
3	4	1	2	2	3	3	19	8	3
7	1	0	2	7	0	4	5	1	14
16	8	9	1	1	2	5	10	2	3

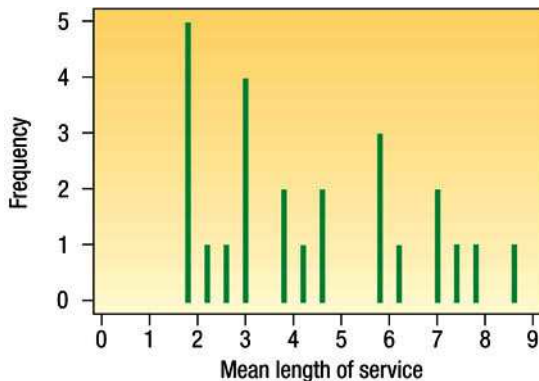
Central Limit Theorem

Distribusi Populasi



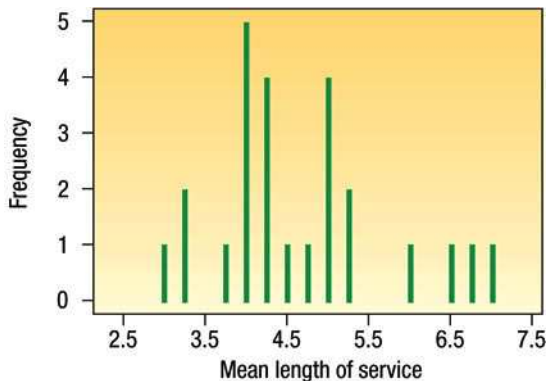
Central Limit Theorem

25 sample masing-masing berukuran $n=5$.



Central Limit Theorem

25 sample masing-masing berukuran $n=20$.



Standard Error dari Mean

Standard Error dari Mean

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \quad (6)$$

Standard Error dari Mean

Standard Error dari Mean

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \quad (7)$$

- 1 Mean dari distribusi sampling dari sample mean akan persis sama dengan mean dari populasi jika kita dapat mengambil semua sample yang mungkin, dengan ukuran yang sama, dari populasi yang diberikan.
- 2 Sebaran dari distribusi sample mean selalu lebih kecil daripada sebaran populasinya. Semakin besar ukuran sample, sebaran dari distribusi sample mean akan semakin mengecil, dengan kata lain standard error dari mean juga akan mengecil.

Penggunaan Distribusi dari Sample Mean: σ diketahui

- 1 Jika distribusi populasi adalah distribusi normal, maka distribusi dari sample mean juga akan berupa distribusi normal.
- 2 Jika distribusi populasi tidak berupa distribusi normal, tetapi $n \geq 30$, maka berdasarkan central limit theorem distribusi sample mean akan berupa distribusi normal.
- 3 Untuk menentukan berapa probabilitas suatu sample mean berada pada wilayah tertentu, maka kita menggunakan:

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad (8)$$

Penggunaan Distribusi dari Sample Mean: σ tidak diketahui

- 1 Jika distribusi populasi adalah distribusi normal, maka distribusi dari sample mean juga akan berupa distribusi normal.
- 2 Jika distribusi populasi tidak berupa distribusi normal, tetapi $n \geq 30$, maka berdasarkan central limit theorem distribusi sample mean akan berupa distribusi normal.
- 3 Untuk menentukan berapa probabilitas suatu sample mean berada pada wilayah tertentu, maka kita menggunakan:

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} \quad (9)$$

Penggunaan Distribusi dari Sample Mean: σ diketahui

Contoh:

Quality control dari sebuah perusahaan minuman mengetahui bahwa jumlah minuman dari setiap botol dapat bervariasi dari satu botol ke botol lainnya. Berdasarkan catatan diketahui bahwa mean jumlah minuman dalam setiap botol adalah 31.2 ounces dengan standard deviasi populasi sebesar 0.4 ounces.

Pada suatu pagi, seorang teknisi mengambil 16 botol secara random. Nilai mean dari 16 botol itu adalah 31.38 ounces. Dengan demikian terdapat selisih (sampling error, $\bar{X} - \mu$) sebesar $31.38 - 31.2 = 0.18$ ounces.

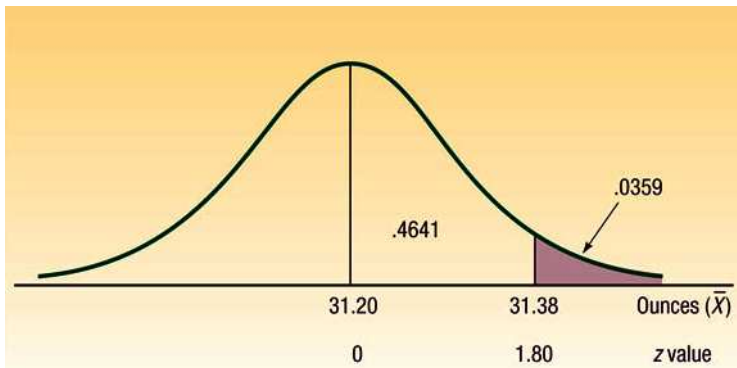
Apakah hasil dari sampling ini merupakan hasil yang masuk akal? Apakah dapat dikatakan bahwa jumlah minuman dalam setiap botol itu terlalu banyak? Dengan kata lain, apakah sampling error sebesar 0.18 ounce itu wajar atau tidak?

Penggunaan Distribusi dari Sample Mean: σ diketahui

Hitung nilai-z:

$$\begin{aligned} z &= \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \\ &= \frac{31.38 - 31.20}{\$0.4/\sqrt{16}} \\ &= 1.80 \end{aligned} \tag{10}$$

Penggunaan Distribusi dari Sample Mean: σ diketahui



Penggunaan Distribusi dari Sample Mean: σ diketahui

Kesimpulan:

Probabilitasnya adalah sebesar 4% bahwa kita akan dapat memilih sample 16 botol dari populasi dengan mean 31.2 ounce dan standard deviasi 0.4 ounce dan memperoleh sample dengan mean 31.38 ounce atau lebih.

Probabilitas 4% adalah cukup kecil. Dengan demikian dapat disimpulkan bahwa prosesnya memasukkan terlalu banyak minuman di dalam botol.