

Nama : Ifham Syafwan Fikri
Github : <https://github.com/syafwan-ux/penugasanPengantarStatistika>
Nomor Mahasiswa : 24/545184/PA/23161
Dosen Pengampu : Faizal Makhrus, S.Kom., M.Sc., Ph.D.

PENUGASAN BOOTSTRAPPING PENGANTAR STATISTIKA

Penugasan Pertama: Frekuensi Distribusi *Sampling* dan *Population*

Berdasarkan permasalahan yang diberikan, sebanyak 26 data dan 24 data dengan 15 data akan dipilih secara random dan diubah menjadi suatu sampel dengan data sampel memanjang hingga ke-500.000 data sampel. Untuk merepresentasikan sedemikian banyak data diperlukan metode sampling. Hasil akan membentuk *normal distribution*, yaitu mean, modus, dan median terletak pada satu titik yang sama di tengah grafik.

Algorithm:

1. Mendeklarasi suatu array dengan data yang telah diberikan
Data Pertama
153, 157, 161, 165, 173, 164, 166, 165, 163, 166, 165, 171, 169, 159, 163, 166, 165, 175, 170, 157, 163, 166, 164, 176, 162, 169;
Data Kedua
156, 156, 156, 157, 158, 159, 160, 160, 161, 161, 161, 162, 163, 164, 165, 167, 169, 169, 169, 170, 172, 172, 173, 175
2. Mengambil 15 data random dari array (tanpa pengembalian) dan mengubahnya menjadi suatu sampel, terurut hingga 500.000 sampel
3. Tiap sampel akan dicari Mean dan Standar Deviasi dan ditempatkan *output*-nya dalam file yang berbeda (hasil dapat dilihat melalui Github)
4. Mean dan Standar Deviasi yang semula ditempatkan dalam file akan dibaca dan dicari tiap Mean dan Standar Deviasinya.
5. Hasil yang didapatkan akan divisualisasikan menggunakan *frequency distribution*.

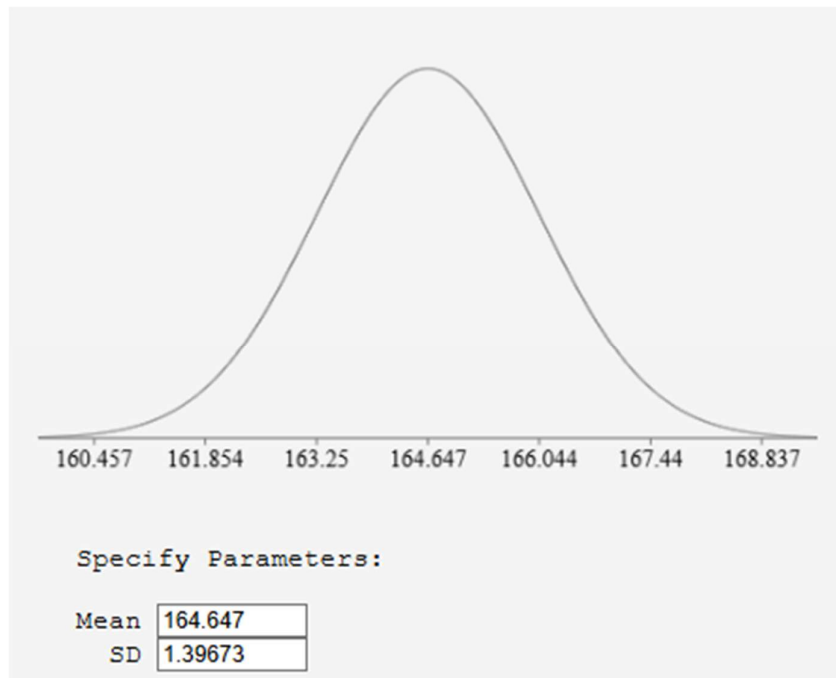
Data Pertama:

153, 157, 161, 165, 173, 164, 166, 165, 163, 166, 165, 171, 169, 159, 163, 166, 165, 175,
170, 157, 163, 166, 164, 176, 162, 169

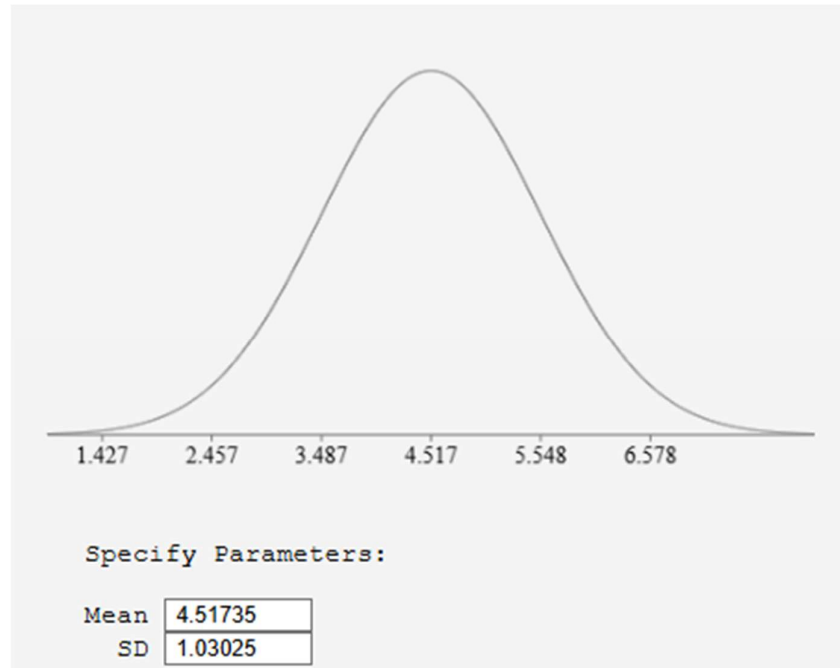
Output

```
./Assignment_1  
Mean dan Standard Deviation dari 500.000 Data Sampel Mean:  
Mean: 164.647  
Standard Deviation: 1.39673  
  
Mean dan Standard Deviation dari 500.000 Data Sampel Std Dev:  
Mean: 4.51735  
Standard Deviation: 1.03025  
Testing complete!° ♦°
```

Graph, normal distribution



Mean Sampling

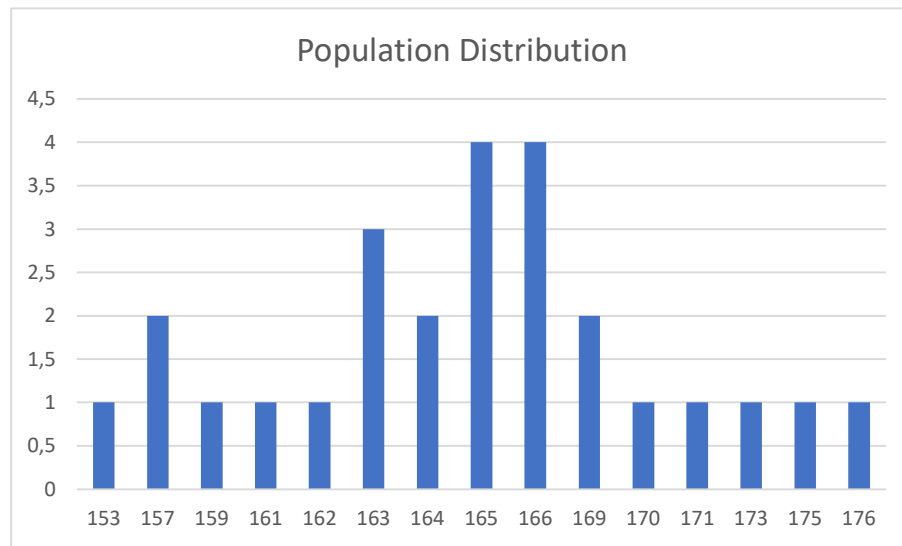


Standard Deviation Sampling

Frequency Table

Apabila menggunakan population distribution, tiap 26 data yang ada akan direpresentasikan satu per satu, baik dalam column maupun line. Karena tiap data memiliki frekuensi yang berbeda-beda dan variabel tinggi badan tidak terlalu banyak, grafik akan tidak terlalu membentuk normal distribution.

Tinggi Badan	Frekuensi
153	1
157	2
159	1
161	1
162	1
163	3
164	2
165	4
166	4
169	2
170	1
171	1
173	1
175	1
176	1



Population Distribution

Data Kedua:

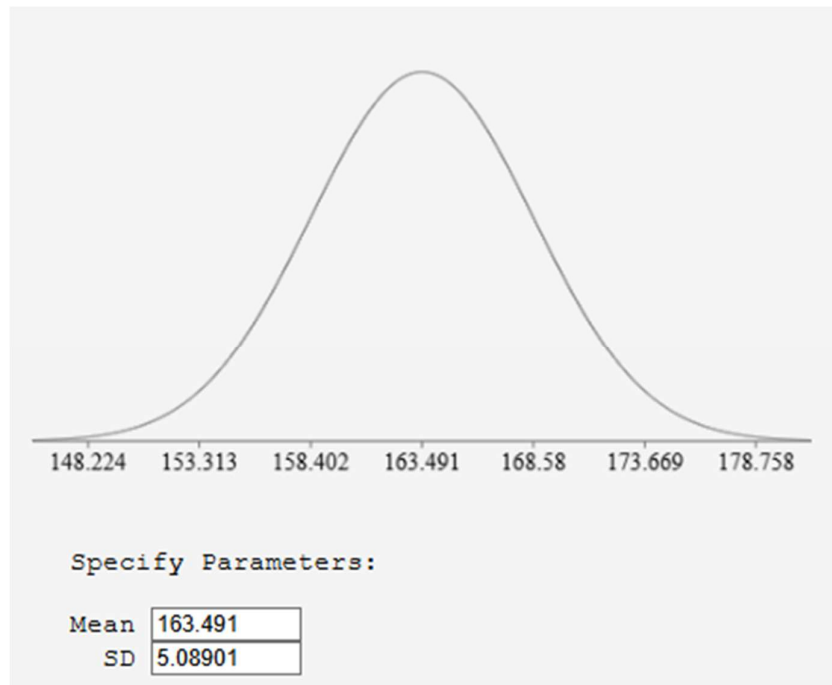
156, 156, 156, 157, 158, 159, 160, 160, 161, 161, 161, 162, 163, 164, 165, 167, 169,
169, 169, 170, 172, 172, 173, 175

Output

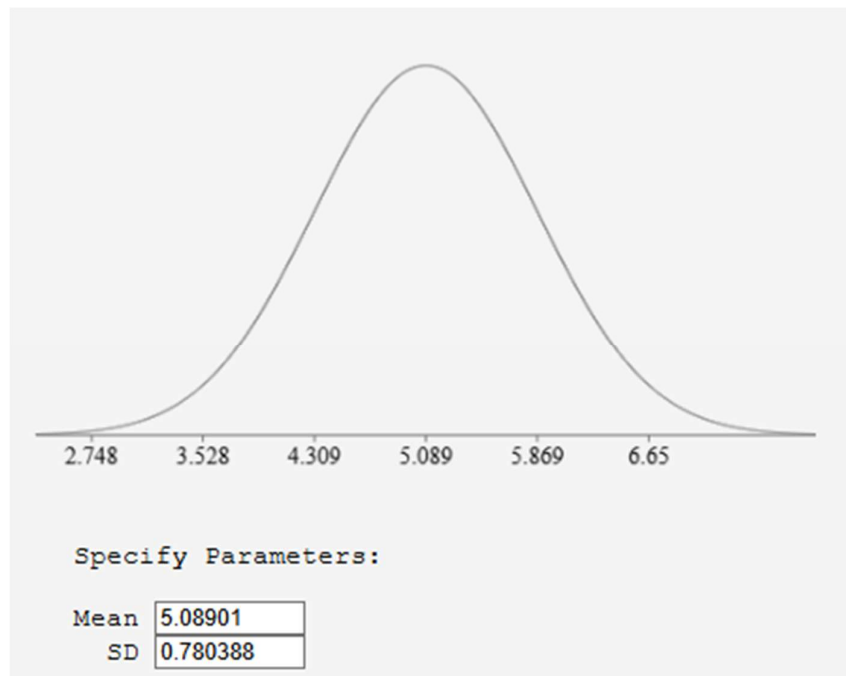
```
Mean dan Standard Deviation dari 500.000 Data Sampel Mean:
Mean: 163.491
Standard Deviation: 1.53125

Mean dan Standard Deviation dari 500.000 Data Sampel Std Dev:
Mean: 5.08901
Standard Deviation: 0.780388
Testing complete!° ◊°
```

Graph, normal distribution



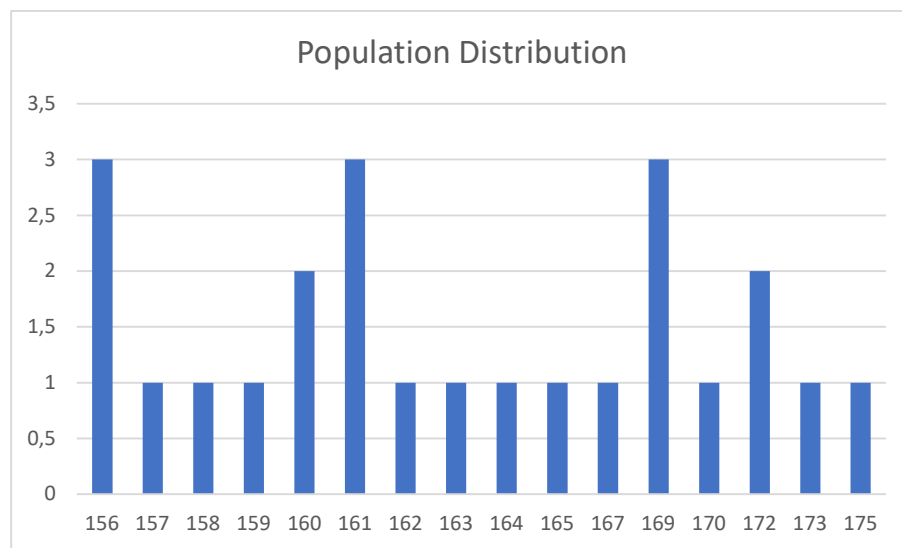
Mean Sampling



Standard Deviation Sampling

Frequency table

Tinggi Badan	Frekuensi
156	3
157	1
158	1
159	1
160	2
161	3
162	1
163	1
164	1
165	1
167	1
169	3
170	1
172	2
173	1
175	1



Population Distribution

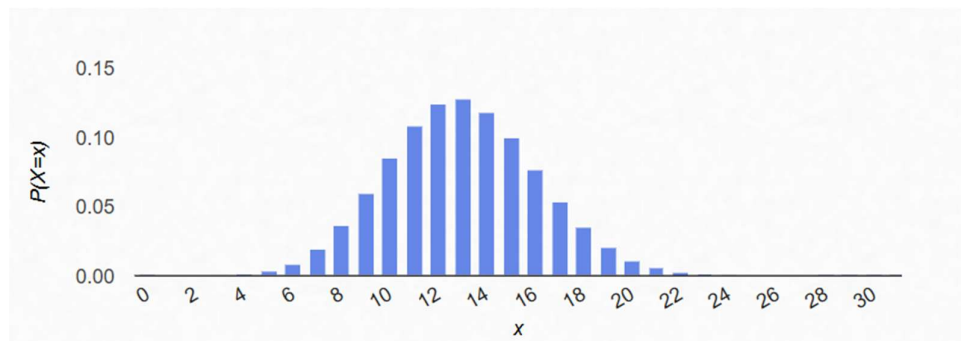
Penugasan Kedua: Peluang Distribusi Binomial

Dalam penerimaan mahasiswa baru S-1 tahun 2024, terdapat 10.678 mahasiswa yang telah dinyatakan sebagai mahasiswa Universitas Gadjah Mada. Baik dari jalur SNBP, SNBT, Mandiri, dan PBUB (IUP tidak termasuk). Dari informasi ini, *Binomial Distribution* dapat dimanfaatkan untuk menghitung probabilitas yang ada. Sebagai contoh, terdapat 2.821 mahasiswa yang masuk melewati jalur SNBP. Berapa peluang yang didapatkan untuk mengambil 50 mahasiswa secara acak dengan tidak lebih 5 di antaranya adalah mahasiswa dari jalur SNBP.

Algorithm:

1. Mencari probabilitas untuk mahasiswa SNBP dari seluruh mahasiswa baru S-1 tahun 2024, kemudian menggunakan konsep komplemen untuk mencari probabilitas sebaliknya
2. Menggunakan Binomial Formula untuk mencari probabilitas tiap kemungkinan. Karena diminta tidak lebih dari 5, kemungkinan yang terjadi adalah total dari $P(X = 0)$ hingga $P(X = 5)$
3. Tiap peluang yang ada dapat direpresentasikan dalam *column graph*

$P(X = 0)$	0.0000002179
$P(X = 1)$	0.0000039125
$P(X = 2)$	0.0000344168
$P(X = 3)$	0.0001977138
$P(X = 4)$	0.0008341057
$P(X = 5)$	0.0027552135
:	:
$P(X = 49)$	0.0000000000
$P(X = 50)$	0.0000000000



Binomial Distribution

Dengan demikian, probabilitas dari pengambilan mahasiswa secara acak dengan tidak lebih dari 5 di antaranya adalah mahasiswa jalur SNBP dapat didapatkan dengan total kumulatif dari tiap kemungkinan yang terkait seperti berikut:

$$\begin{aligned}
 P(X \leq 5) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) \\
 &= 0.0000002179 + 0.0000039125 + 0.0000344168 + 0.0001977138 + 0.0008341057 + \\
 &\quad 0.0027552135 \\
 &= \mathbf{0.0038255802}
 \end{aligned}$$

Penugasan Ketiga: *Standard normal distribution*

Untuk mencari total probabilitas dari *binomial distribution* menggunakan *standard normal distribution*, kita dapat menggunakan sampling dan menghitung area terkait dengan Z-Score. Berikut merupakan persamaan yang dipakai untuk mengkonversi nilai yang didapatkan dari distribusi binomial:

- Mean adalah probabilitas sukses
- X variable menjadi p sample
- Standard Deviation adalah $\sqrt{pq/n}$
- Jadinya Z adalah hasil dari $(p \text{ sample} - p) / \text{Standard Deviation}$

Apabila kita mencoba untuk mencari $P(X \leq 5)$, kita dapat mengubahnya ke dalam Z-Score agar lebih mudah untuk dihitung probabilitasnya dan mencari areanya. Sebagai contoh:

$$X = (0.1 - 0.26) / [((0.26)(0.74)/50)]^{(1/2)}$$

$$X = (-0.16) / (0.06)$$

$$X = -2.67$$

Melihat nilai yang didapatkan, yaitu -2,67, kita akan mendapatkan nilai 0.0038. Nilai ini hampir sama dengan total kumulatif probabilitas dari Penugasan Kedua. Tabel Z-Score dan grafik dapat dilihat sebagai berikut:

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



Results:
Area (probability) =