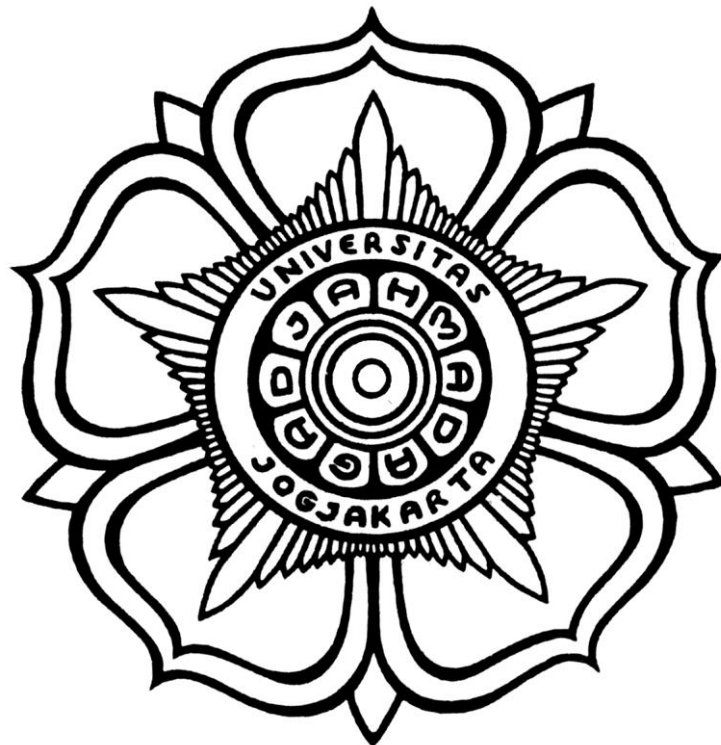


**LAPORAN TUGAS FINAL PENGANTAR STATISTIKA**  
**INTERVAL KEPERCAYAAN USIA RATA-RATA PEMAIN SEPAKBOLA**  
**DI LIGA INGGRIS: ANALISIS MENGGUNAKAN METODE**  
**BOOTSTRAP SAMPLING DAN DISTRIBUSI NORMAL Z-SCORE**



**DISUSUN OLEH :**  
**TENGKU SYAID FARHAN**  
**(22/493998/PA/21234)**  
**PROGRAM ILMU KOMPUTER**  
**FMIPA**

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar belakang**

Dalam analisis statistik, uji interval kepercayaan merupakan metode yang digunakan untuk mengestimasi rentang nilai yang memungkinkan untuk mengandung parameter populasi dengan tingkat kepercayaan tertentu. Dalam kasus ini, kita akan fokus pada uji interval kepercayaan untuk usia rata-rata pemain sepakbola di Liga Inggris.

Metode bootstrap sampling dan distribusi normal Z-score adalah dua metode yang dapat digunakan untuk menghitung interval kepercayaan dalam analisis ini. Metode bootstrap sampling memungkinkan kita untuk menghasilkan banyak sampel bootstrap dari data sampel yang tersedia, sedangkan distribusi normal Z-score memungkinkan kita untuk menggunakan nilai-nilai dalam distribusi normal standar untuk menghitung interval kepercayaan.

### **1.2 Tujuan percobaan**

Tujuan dari analisis ini adalah untuk mengestimasi interval kepercayaan usia rata-rata pemain sepakbola di Liga Inggris menggunakan metode bootstrap sampling dan distribusi normal Z-score. Dengan melakukan ini, kita dapat membuat perkiraan rentang nilai yang memungkinkan untuk mengandung usia rata-rata populasi dengan tingkat kepercayaan tertentu.

### **1.3 Manfaat percobaan**

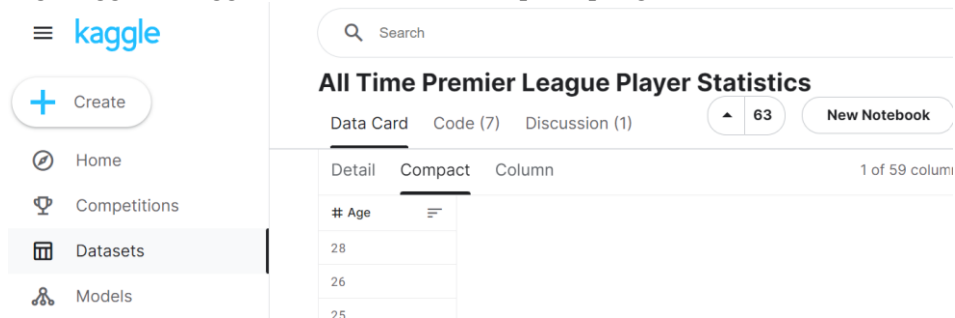
Manfaat dari analisis ini adalah memberikan informasi yang lebih valid dan reliabel tentang usia rata-rata pemain sepakbola di Liga Inggris. Dengan menggunakan metode bootstrap sampling dan distribusi normal Z-score, kita dapat menghasilkan interval kepercayaan yang memberikan estimasi yang lebih akurat dan mengukur tingkat ketidakpastian dalam estimasi tersebut.

## BAB II

### ISI

#### 2.1 Materi dan metode

Dalam penelitian ini, akan menguji interval kepercayaan usia rata-rata pemain sepakbola di Liga Inggris menggunakan metode bootstrap sampling dan distribusi normal Z-score.



#	Age
28	
26	
25	

1. Bootstrap sampling digunakan ketika sampel acak dari data tersedia dan estimasi interval kepercayaan untuk parameter statistik diinginkan. Metode ini melibatkan pembuatan banyak sampel bootstrap dengan pengambilan sampel acak dengan penggantian dari data asli. Langkah-langkahnya meliputi pengambilan sampel acak dengan penggantian, perhitungan statistik dari sampel bootstrap, dan pengulangan untuk memperoleh banyak sampel bootstrap  $p$ . Dalam metode ini, interval kepercayaan dapat dihitung menggunakan metode percentile.
2. Distribusi Z digunakan ketika data yang tersedia cukup besar dan asumsi bahwa data terdistribusi secara normal dapat dipertahankan. Metode ini melibatkan perhitungan statistik dari sampel, seperti rata-rata atau proporsi, dan perhitungan deviasi standar dari sampel. Interval kepercayaan kemudian dihitung menggunakan rumus yang sesuai dengan distribusi Z, seperti rumus Z untuk interval kepercayaan proporsi atau rumus Z untuk interval kepercayaan rata-rata

## 2.2 Tinjauan

### a. Metode Bootstrap Sampling

Implementasi menggunakan source code python

```
1  import random
2  import pandas as pd
3  import numpy as np
4  import matplotlib.pyplot as plt
5  import openpyxl
6
7  data = pd.read_csv('dataset.csv')
8  m = 500000
9  sample = []
10 sample = data['Age'].sample(n=100, random_state=42).tolist()
11 Size_Sample = len(sample)
```

mengimpor beberapa library seperti random, pandas, numpy, matplotlib.pyplot, dan openpyxl. Selanjutnya, kode membaca file CSV yang disimpan dalam variabel data dan mendefinisikan variabel m dengan nilai 500,000. Kemudian, kode mengambil sampel acak sebanyak 100 data dari kolom 'Age' dalam DataFrame data dan menyimpannya dalam variabel sample. Terakhir, kode menghitung panjang sample dan menyimpan hasilnya dalam variabel Size\_Sample.

```
1  Stat_Sample=[]
2  for _ in range(m):
3      Sample_Boots = random.choices(sample, k=Size_Sample)
4      rata_Sample = np.mean(Sample_Boots)
5      Stat_Sample.append(rata_Sample)
6
7  print('Sampling 1 : ')
8  print(sample)
9  Stat_Sample.sort()
```

perulangan sebanyak m kali. Setiap iterasi perulangan, dilakukan pengambilan sampel acak sebanyak Size\_Sample dari sample menggunakan fungsi random.choices(). Rata-rata

dari sampel tersebut dihitung menggunakan fungsi `np.mean()` dan disimpan dalam variabel `rata_Sample`. Hasil rata-rata tersebut kemudian ditambahkan ke dalam daftar `Stat_Sample`. Setelah perulangan selesai, hasil sampling pertama dicetak menggunakan `print()`, dan `Stat_Sample` diurutkan menggunakan metode `sort()`.

Output :

```
Sampling 1 :  
[28.0, 20.0, 26.0, 22.0, 20.0, 20.0, 22.0, 32.0, 21.0, 20.0, 29.0, 32.0, 18.0, 26.0, 27.0,  
20.0, 23.0, 22.0, 22.0, 28.0, 24.0, 27.0, 18.0, 32.0, 34.0, 32.0, 24.0, 31.0, 17.0, 31.0,  
23.0, 26.0, 29.0, 27.0, 21.0, 22.0, 33.0, 27.0, 32.0, 22.0, 31.0, 30.0, 32.0, 31.0, 29.0,  
30.0, 26.0, 23.0, 29.0, 27.0, 26.0, 34.0, 22.0, 31.0, 24.0, 26.0, 20.0, 19.0, 25.0, 29.0,  
22.0, 25.0, 20.0, 24.0, 19.0, 19.0, 29.0, 29.0, 29.0, 35.0, 26.0, 27.0, 28.0, 26.0, 33.0,  
33.0, 22.0, 25.0, 19.0, 28.0, 24.0, 30.0, 27.0, 29.0, 32.0, 21.0, 25.0, 22.0, 23.0, 34.0,  
25.0, 20.0, 27.0, 21.0, 29.0, 20.0, 29.0, 23.0, 32.0, 21.0]
```

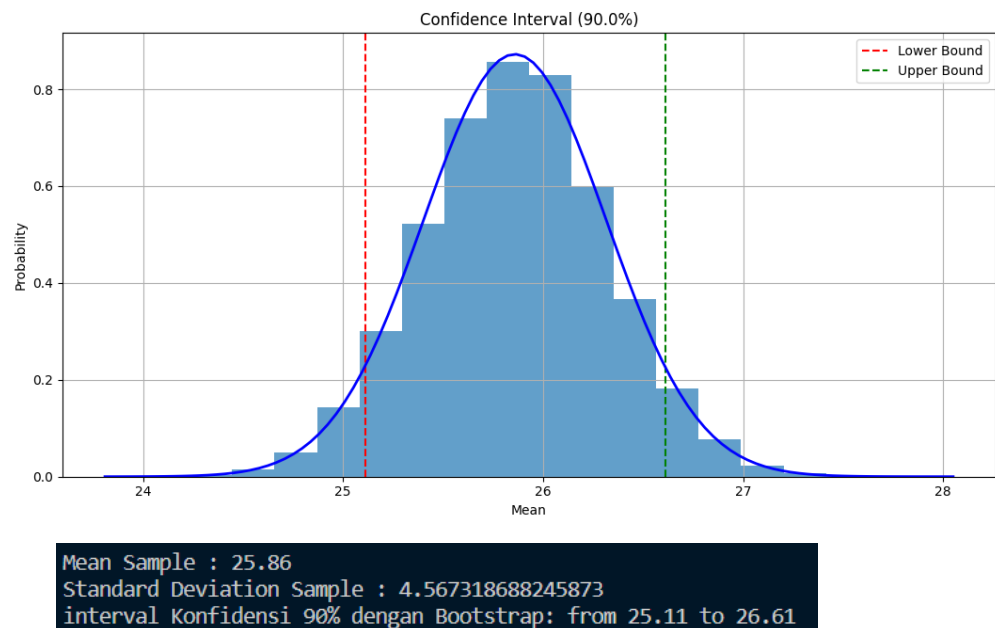
```
1 confidence_level = 0.90  
2 confidence_interval = pd.Series(Stat_Sample).quantile([0.5 - confidence_level/2, 0.5 + confidence_level/2])  
3 lower_bound, upper_bound = confidence_interval.values
```

menghitung interval kepercayaan dari data yang telah disimulasikan sebelumnya. Tingkat kepercayaan ditentukan sebagai 90%. Menggunakan metode `quantile()` dari objek `Series` yang dibuat dari `Stat_Sample`, kita menghitung kuantil pada titik 0,5 dikurangi setengah dari tingkat kepercayaan yang dibagi 2, dan kuantil pada titik 0,5 ditambah setengah dari tingkat kepercayaan yang dibagi 2. Hasilnya adalah dua nilai yang mewakili batas bawah (`lower_bound`) dan batas atas (`upper_bound`) dari interval kepercayaan.

```
1 plt.show()  
2 print(" ")  
3 print(f"Mean Sample : {np.mean(sample)}")  
4 print(f"Standard Deviation Sample : {np.std(sample)}")  
5 print(f"interval Konfidensi 90% dengan Bootstrap: from {lower_bound} to {upper_bound}")
```

menggunakan `plt.show()` untuk menampilkan grafik yang telah dibuat. Kemudian, kita mencetak rata-rata (mean) dari sampel menggunakan `np.mean(sample)`, simpangan baku (standard deviation) menggunakan `np.std(sample)`, dan interval kepercayaan 90% hasil simulasi bootstrap yang telah dihitung sebelumnya. Nilai-nilai tersebut dicetak menggunakan `print()`.

Output :



Output berupa mean sample bernilai 25.86, standar deviasi 4.567318688245873, dan confidence interval 25.11 – 26.61.

#### b. Metode Distribusi Z

Metode distribusi Z digunakan ketika data terdistribusi secara normal dan ukuran sampel cukup besar.

##### Contoh Kasus diambil dari data sample

suatu pengamat sepakbola ingin mengetahui berapa rata-rata umur dari pemain sepak bola pada liga inggris, diadakan sampling 100 sample dan didapat mean = 25.860409 standard deviasi = 4.567318688245873 tentukan interval konfidensi 90 % !

Langkah- Langkah :

1. Interval estimasi

$$\bar{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$A < \mu < B$$

2. Z score menggunakan distribusi Z

$$z_{(0,05)} = 1,64$$

3. Menghitung interval konfidensi dan margin of error

$$s = 4.567318688245873$$

$$n = 100$$

$$\bar{x} = 25.860409$$

$$\begin{aligned} A &= \bar{x} - Z_{(10\%/2)} \times \frac{s}{\sqrt{n}} \\ &= 25.860409 - Z_{(0,05)} \times \frac{4.567318688245873}{\sqrt{100}} \\ &= 25.860409 - 1,64 \times \frac{4.567318688245873}{\sqrt{100}} \\ &= 25.111 \end{aligned}$$

$$\begin{aligned} B &= \bar{x} + Z_{(10\%/2)} \times \frac{s}{\sqrt{n}} \\ &= 25.860409 + Z_{(0,05)} \times \frac{4.567318688245873}{\sqrt{100}} \\ &= 25.860409 + 1,64 \times \frac{4.567318688245873}{\sqrt{100}} \\ &= 26.61 \end{aligned}$$

4. Didapatkan Interval konfidensi 90%

$$25.11 < \mu < 26.61$$

## **BAB II**

### **HASIL DAN ANALISIS**

Setelah menerapkan kedua metode, dapat dibandingkan hasil interval kepercayaan yang diperoleh.

Hasil dari metode bootstrap sampling menunjukkan rata-rata sampel sebesar 25,86 dengan simpangan baku sebesar 4,57. Interval kepercayaan 90% yang dihasilkan adalah 25,11 - 26,61.

Sementara itu, hasil dari metode distribusi Z juga menunjukkan interval kepercayaan 90% yang serupa, dengan rata-rata sampel sebesar 25,86 dan simpangan baku sebesar 4,57. Interval kepercayaan yang dihasilkan adalah 25,11 - 26,61.

Berdasarkan hasil ini, kedua metode memberikan hasil yang serupa atau identik dalam mengestimasi interval kepercayaan usia rata-rata pemain sepakbola di Liga Inggris. Kedua metode dapat digunakan dengan keyakinan untuk memberikan estimasi yang akurat serta mengukur tingkat ketidakpastian dalam estimasi tersebut.



## **BAB IV**

### **SIMPULAN**

Dalam analisis ini, digunakan metode bootstrap sampling dan distribusi normal Z-score untuk mengestimasi interval kepercayaan usia rata-rata pemain sepakbola di Liga Inggris. Kedua metode tersebut menghasilkan hasil yang serupa, dengan interval kepercayaan yang hampir identik.

Metode bootstrap sampling memungkinkan dilakukannya pembuatan banyak sampel bootstrap dari data sampel yang tersedia, sementara distribusi normal Z-score memanfaatkan asumsi bahwa data terdistribusi secara normal dan ukuran sampel yang cukup besar.

Dalam kedua metode tersebut, tingkat kepercayaan tertentu (misalnya, 90%) dapat diukur untuk menghasilkan interval kepercayaan yang mungkin mengandung usia rata-rata populasi dengan tingkat kepercayaan tersebut.

Kedua metode ini memberikan estimasi yang akurat serta memberikan informasi mengenai tingkat ketidakpastian dalam estimasi tersebut. Oleh karena itu, kedua metode ini dapat digunakan untuk memberikan informasi yang lebih valid dan reliabel mengenai usia rata-rata pemain sepakbola di Liga Inggris.