Imports

In [2]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

from scipy.stats import norm
```

CLT

Population

In [3]:
```
df = pd.read_csv("weight-height.csv")
df
```

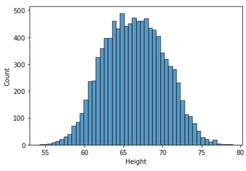
Out[3]:

	Gender	Height	Weight
0	Male	73.847017	241.893563
1	Male	68.781904	162.310473
2	Male	74.110105	212.740856
3	Male	71.730978	220.042470
4	Male	69.881796	206.349801
...	...	...	...
9995	Female	66.172652	136.777454
9996	Female	67.067155	170.867906
9997	Female	63.867992	128.475319
9998	Female	69.034243	163.852461
9999	Female	61.944246	113.649103

10000 rows × 3 columns

In [72]:
```
sns.histplot(df["Height"])
```

Out[72]:
```
<AxesSubplot:xlabel='Height', ylabel='Count'>
```



In [18]:
```
df["Height"].mean()
```

Out[18]:
```
66.36755975482106
```

In [78]:
```
sigma=df["Height"].std()
sigma
```

Out[78]:
```
3.8475281207732324
```

Sample of 5

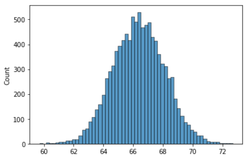
In [16]:
```
df["Height"].sample(5)
```

Out[16]:
```
5299    59.304007
2895    71.382577
8735    61.426205
1596    66.628435
1482    68.610922
Name: Height, dtype: float64
```

In [65]:
```
sample_5_mean=[np.mean(df["Height"].sample(5)) for i in range(10000)]
# sample_5_mean
```

In [66]:
```
sns.histplot(sample_5_mean)
```

Out[66]:
```
<AxesSubplot:ylabel='Count'>
```



In [67]:
```
np.mean(sample_5_mean)
```

Out[67]:
```
66.37516623229443
```

In [68]:
```
np.std(sample_5_mean)
```

Out[68]:
```
1.7295648087604485
```

In [79]:
```
sigma/np.sqrt(5)
```

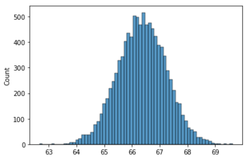
Out[79]:
```
1.7206668846781936
```

Sample of 20

In [69]:
```
sample_20_mean=[np.mean(df["Height"].sample(20)) for i in range(10000)]
# sample_20_mean
```

In [71]:
```
sns.histplot(sample_20_mean)
```

Out[71]:
```
<AxesSubplot:ylabel='Count'>
```



In [73]:
```
np.mean(sample_20_mean)
```

Out[73]:
```
66.36291592254688
```

In [74]:
```
np.std(sample_20_mean)
```

Out[74]:
```
0.8650732838955196
```

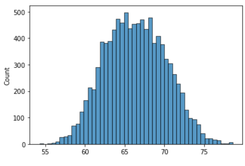
In [80]:
```
sigma/np.sqrt(20)
```

Out[80]:
```
0.8603334423390968
```

In [75]:
```
sample_20_mean=[np.mean(df["Height"].sample(1)) for i in range(10000)]
# sample_20_mean
```

In [76]:
```
sns.histplot(sample_20_mean)
```

Out[76]:
```
<AxesSubplot:ylabel='Count'>
```



Q1. Blood Pressure Problem Statement

Systolic blood pressure of a group of people is known to have an average of 122 mmHg and a standard deviation of 10 mmHg

Calculate the probability that the average blood pressure of 16 people will be greater than 125 mmHg.

In [82]:
```
1-norm.cdf(1.2)
```

Out[82]:
```
0.11506967022170822
```

In [83]:
```
1-norm.cdf(x=125,loc=122,scale=2.5)
```

Out[83]:
```
0.11506967022170822
```

Q2. Weekly Tooth Paste Sales Problem Statement

Weekly toothpaste sales have a mean 1000 and std dev 200. What is the probability that the average weekly sales next month is more than 1110?

In [85]:
```
1-norm.cdf((1110-1000)/100)
```

Out[85]:
```
0.13566606094638267
```

In [84]:
```
1-norm.cdf(x=1110,loc=1000,scale=100)
```

Out[84]:
```
0.13566606094638267
```

Q3. Ecommerce Problem Statement

In an e-commerce website, the average purchase amount per customer is $80 with a standard deviation of $15. If we randomly select a sample of 50 customers, what is the probability that the average purchase amount in the sample will be less than $75?

In [87]:
```
norm.cdf(x=75,loc=80,scale=(15/np.sqrt(50))) # less than $75
```

Out[87]:
```
0.009211062727049501
```

In [89]:
```
norm.cdf((75-80)/(15/np.sqrt(50)))
```

Out[89]:
```
0.009211062727049501
```

In [90]:
```
1- norm.cdf(x=75,loc=80,scale=(15/np.sqrt(50))) # more than $75
```

Out[90]:
```
0.9907889372729505
```

In [ ]:

In [ ]: