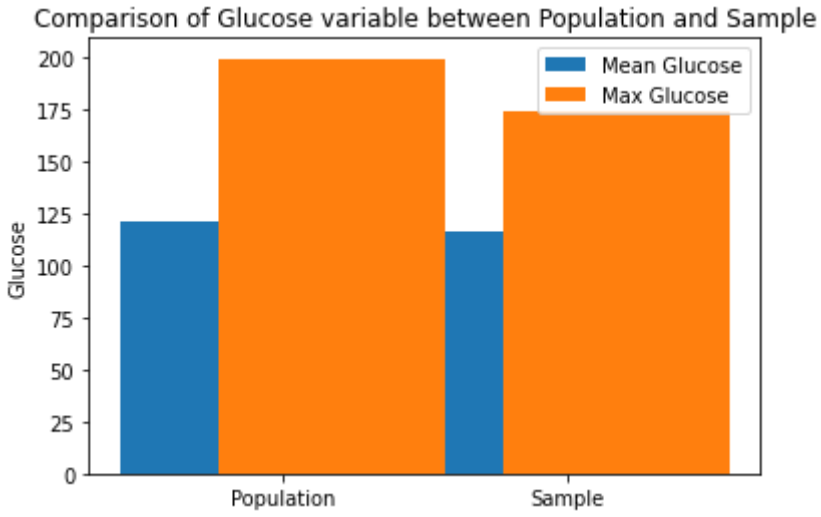
In [43]:

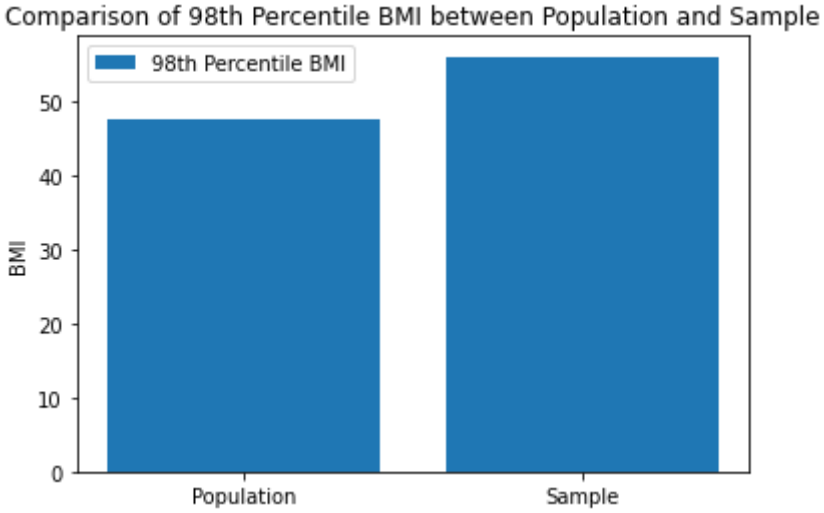
```
import pandas as pd
import numpy as np
import random
import matplotlib.pyplot as plt


np.random.seed(16336381)

datadiab = pd.read_csv("C:/Users/91938/Desktop/diabetes.csv")

sample = datadiab.sample(n=25)
sample_mean_glucose = sample["Glucose"].mean()
sample_max_glucose = sample["Glucose"].max()

population_mean_glucose = datadiab["Glucose"].mean()
population_max_glucose = datadiab["Glucose"].max()



x = np.arange(2)  # the label locations
width = 0.35

fig, ax = plt.subplots()

me = ax.bar(x, [population_mean_glucose, sample_mean_glucose], label="Mean Glucose")
mx = ax.bar(x+width, [population_max_glucose, sample_max_glucose], label="Max Glucos




ax.set_ylabel("Glucose")
ax.set_title("Comparison of Glucose variable between Population and Sample")
ax.set_xticks(x + width/2)
ax.set_xticklabels(["Population","Sample"])

ax.legend()
plt.show()
```



In [45]:

```
sample_98_percentile = np.percentile(datadiab["BMI"].sample(n=25), 98)
population_98_percentile = np.percentile(datadiab["BMI"], 98)

fig, ax = plt.subplots()
ax.bar(["Population", "Sample"], [population_98_percentile, sample_98_percentile], l
ax.set_ylabel("BMI")
```

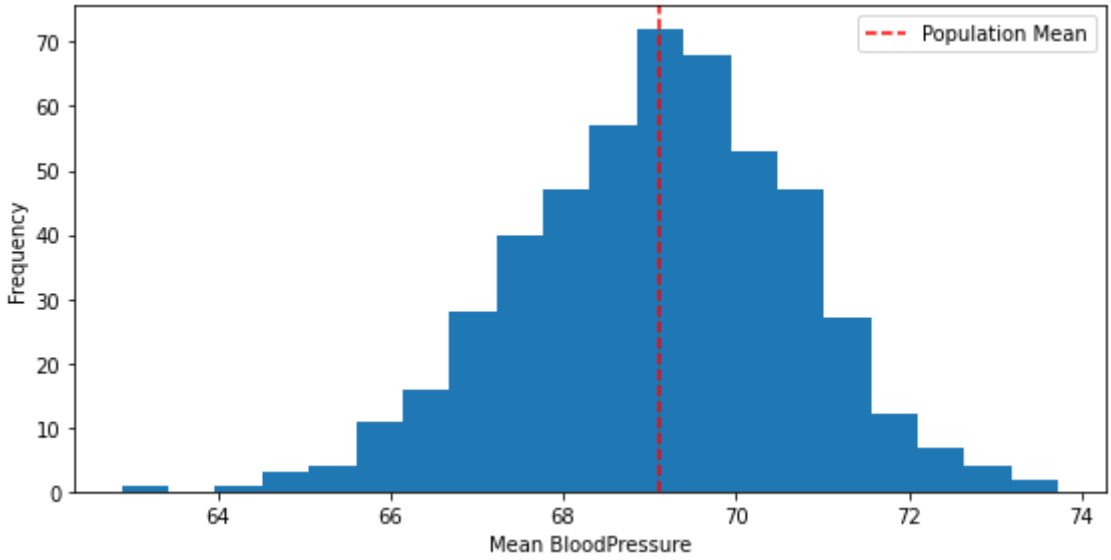
```
ax.set_title("Comparison of 98th Percentile BMI between Population and Sample")
ax.legend()
plt.show()
```

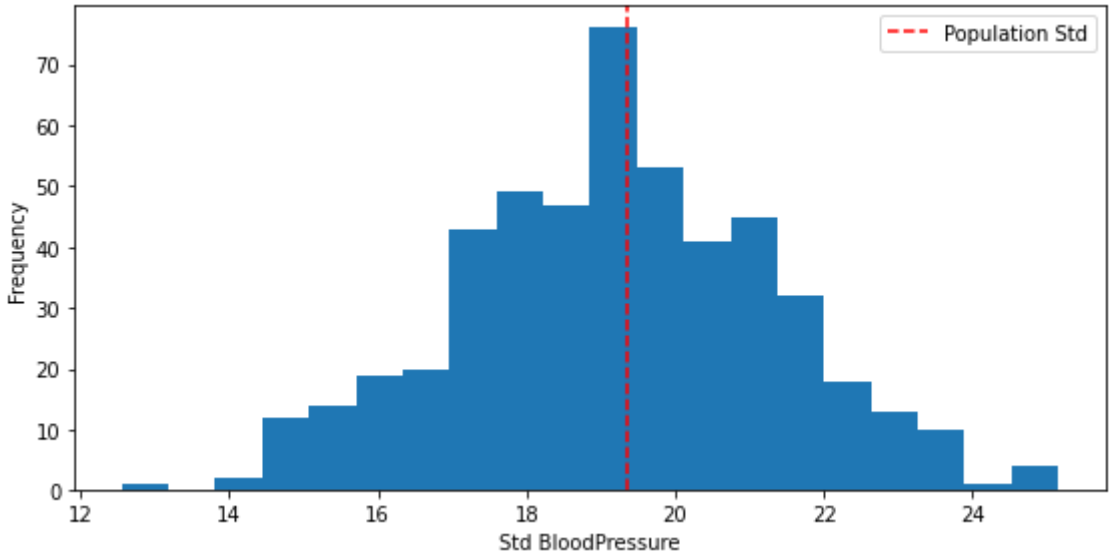


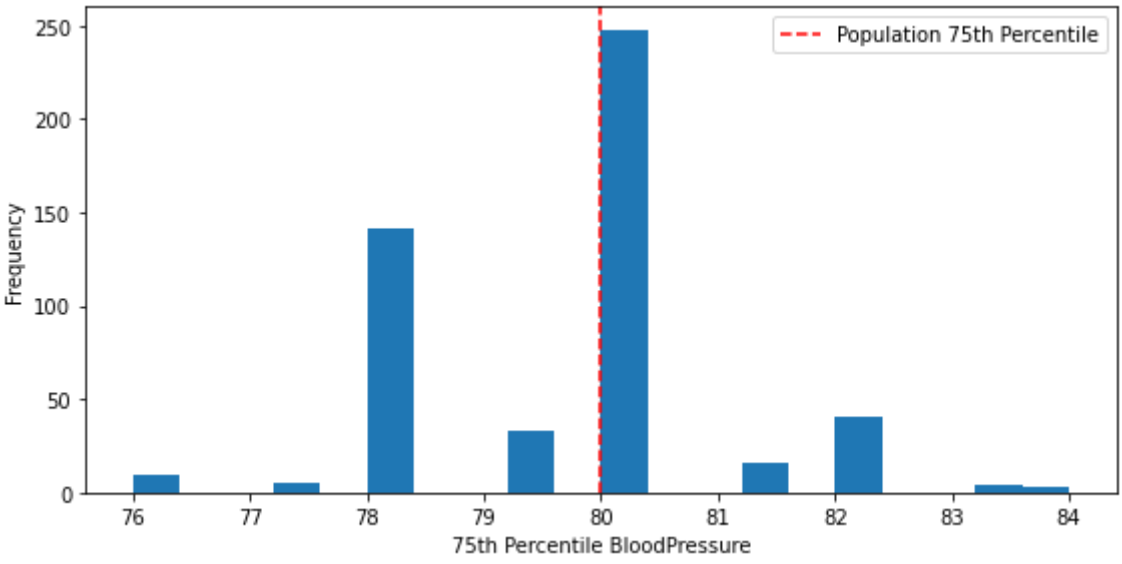
In [48]:

```
n_samples = 500
sample_size = 150

sample_means = np.zeros(n_samples)
sample_stds = np.zeros(n_samples)
sample_percentiles = np.zeros(n_samples)

def bootstrap_sample(datadiab):
    return datadiab.sample(n=len(datadiab), replace=True)

for i in range(n_samples):
    sample = datadiab["BloodPressure"].sample(n=sample_size, replace=True)
    sample_means[i] = sample.mean()
    sample_stds[i] = sample.std()
    sample_percentiles[i] = np.percentile(sample, 75)

population_mean_bp = datadiab["BloodPressure"].mean()
population_std_bp = datadiab["BloodPressure"].std()
population_percentile_bp = np.percentile(datadiab["BloodPressure"], 75)

fig, axes = plt.subplots(3, 1, figsize=(8, 12))

# vals = [{
#     'idx':0,

# }]

# for i in range(3):

axes[0].hist(sample_means, bins=20)
axes[0].axvline(population_mean_bp, color="red", linestyle="--", label="Population M
axes[0].set_xlabel("Mean BloodPressure")
axes[0].set_ylabel("Frequency")
axes[0].legend()

axes[1].hist(sample_stds, bins=20)
axes[1].axvline(population_std_bp, color="red", linestyle="--", label="Population St
axes[1].set_xlabel("Std BloodPressure")
axes[1].set_ylabel("Frequency")
axes[1].legend()

axes[2].hist(sample_percentiles, bins=20)
```

```
axes[2].axvline(population_percentile_bp, color="red", linestyle="--", label="Popula
axes[2].set_xlabel("75th Percentile BloodPressure")
axes[2].set_ylabel("Frequency")
axes[2].legend()

plt.tight_layout()
plt.show()
```







In [ ]: