

Edelstein_EvanGraded_Lab1

March 11, 2021

```
[6]: #import pandas, numpy, matplotlib lib and previous modules from practise labs
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import sys
sys.path.insert(0, '/Users/evanedelstein/Desktop/School/2021/Spring2021/
↳BTM-6000/Mods')
from Biostats import *
from numpy import sqrt
# read dataset
df = pd.read_csv("/Users/evanedelstein/Desktop/School/2021/Spring2021/BTM-6000/
↳Module7/Module 7 - Graded Lab 1 - NHANES subset.csv")
```

```
[4]: # describe data set
print(df["SBP"].describe())
std = np.std(df["SBP"])
SE = std / sqrt(1000)
print("Standard error:",SE)
```

```
count      1000.000000
mean        124.922000
std          20.284313
min           82.000000
25%          112.000000
50%          122.000000
75%          136.000000
max          230.000000
Name: SBP, dtype: float64
Standard error: 0.6411255072136812
```

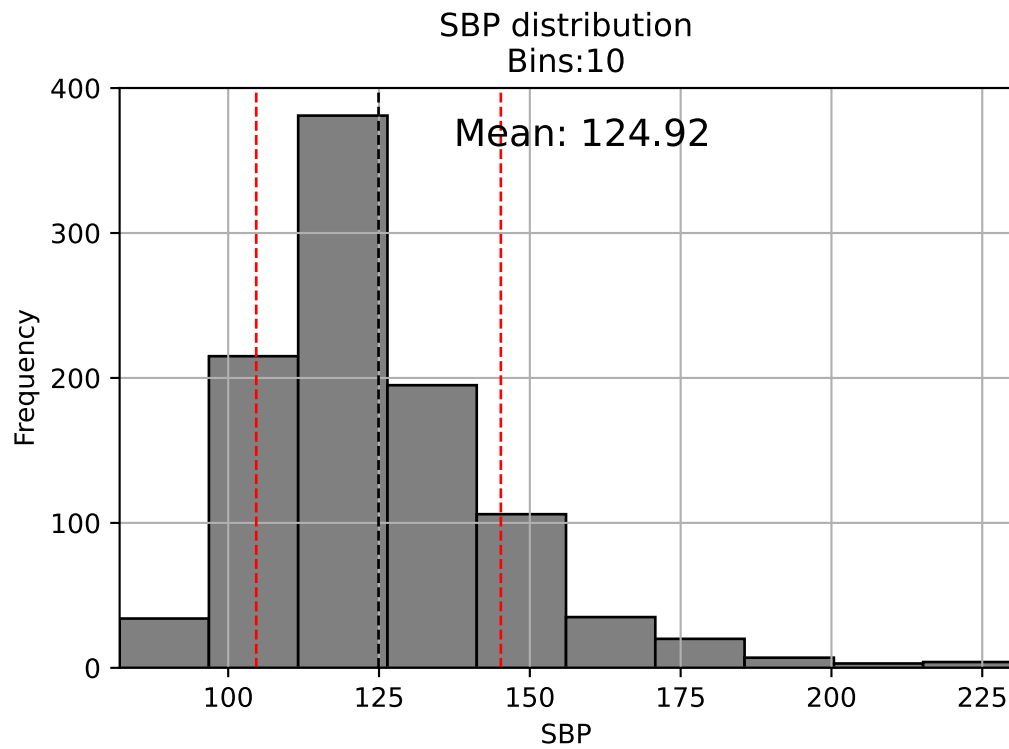
```
[3]: # output IQR
IQR(df, "SBP")
```

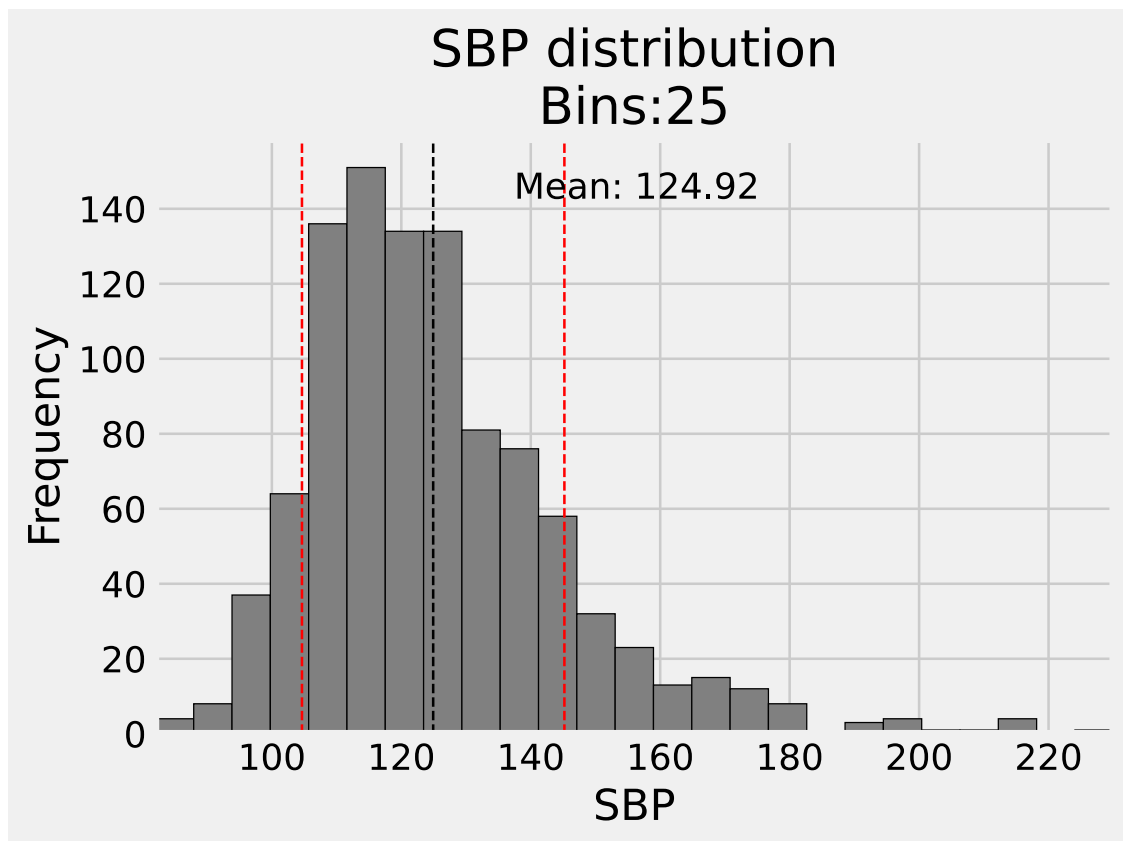
IQR: 24.0

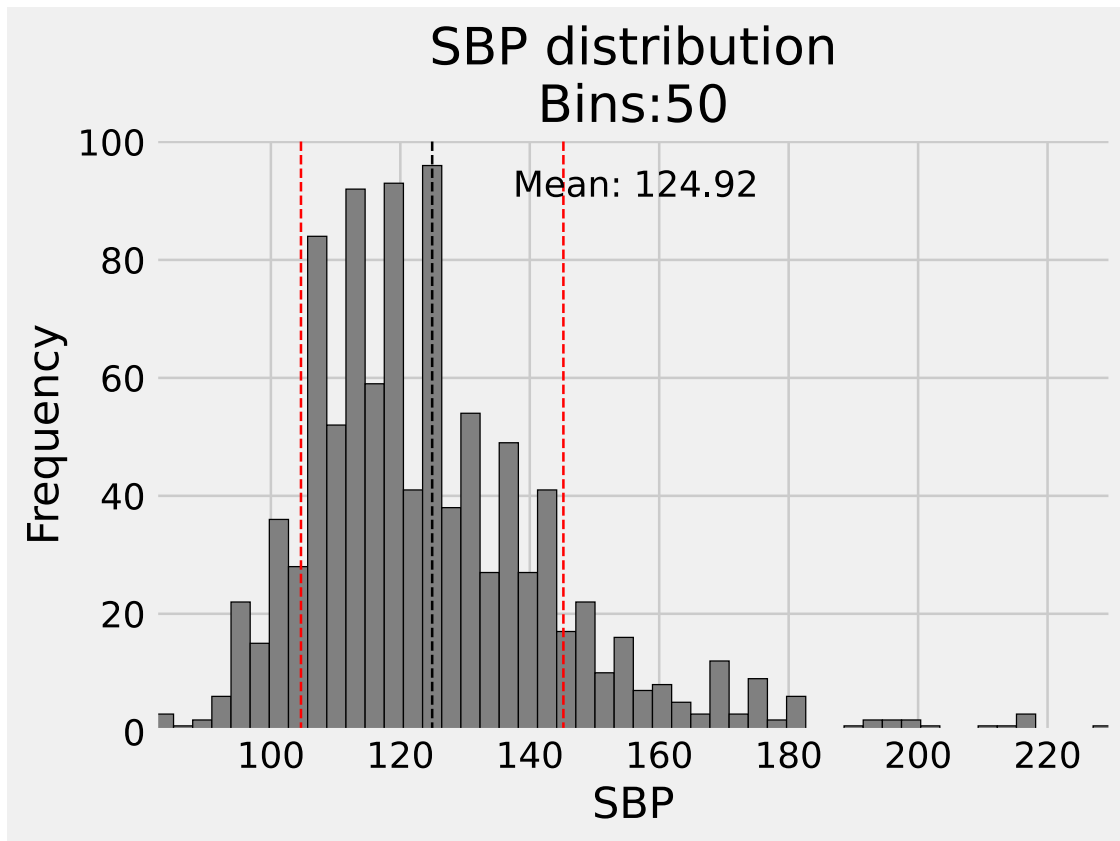
```
[5]: # plot 3 histplots with different bin sizes
hist_plot(df,"SBP",False,10)
hist_plot(df,"SBP",False,25)
```

```
hist_plot(df, "SBP", False, 50)
```

```
[5]: <module 'matplotlib.pyplot' from  
'/Users/evanedelstein/.pyenv/versions/RajiResearch3.9.1/lib/python3.9/site-  
packages/matplotlib/pyplot.py'>
```







```
[14]: # print standev and stdev
std = np.std(df["SBP"])
```

```
SE = std / sqrt(1000)
print("std:",std)
print("Standard error:",SE)
print()
```

```
std: 20.27416868825945
Standard error: 0.6411255072136812
```

```
[10]: # create three stes of 100 samples and output mean, stddev, sterror and histplt
```

```
# for i in range(1,4):
#     print("sample:",i)
#     sample = df.sample(n=100)
#     mean = np.mean(sample["SBP"])
#     std = np.std(sample["SBP"])
#     SE = std / sqrt(100)
#     print("mean:",mean)
#     print("std:",std)
```

```

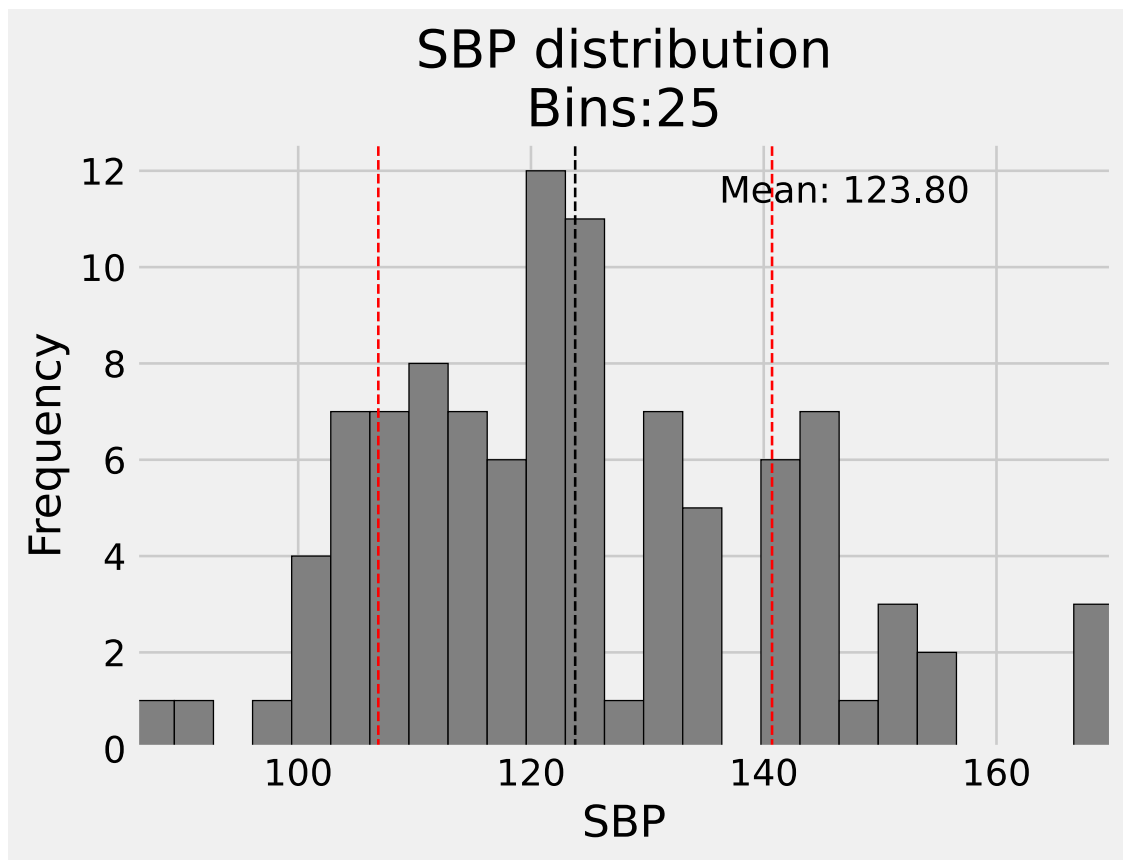
#     print("Standard error:",SE)
#     hist_plot(sample,"SBP",False,15)
from matplotlib.backends.backend_pdf import PdfPages
with PdfPages("SBP_samples_hist.pdf") as pdf:
    for i in range(1,4):
        print("sample:",i)
        sample = df.sample(n=100)
        mean = np.mean(sample["SBP"])
        std = np.std(sample["SBP"])
        SE = std / sqrt(100)
        print("mean:",mean)
        print("std:",std)
        print("Standard error:",SE)
        plt.figure(figsize=(3, 3))
        plot = hist_plot(sample,"SBP",False,25)
        pdf.savefig()
plt.close()

```

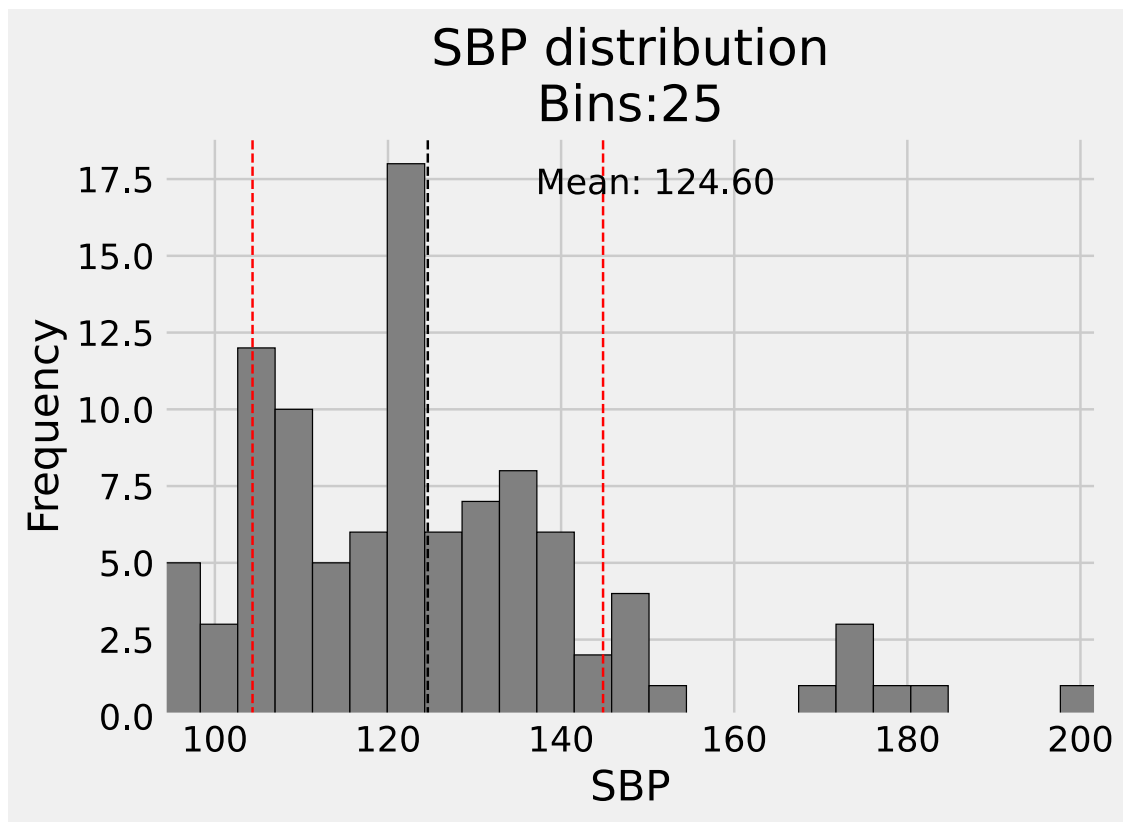
```

sample: 1
mean: 123.8
std: 16.915082027587097
Standard error: 1.6915082027587096
sample: 2
mean: 124.6
std: 20.247468977627793
Standard error: 2.024746897762779
sample: 3
mean: 125.06
std: 20.219703261917573
Standard error: 2.021970326191757
<Figure size 216x216 with 0 Axes>

```



<Figure size 216x216 with 0 Axes>



<Figure size 216x216 with 0 Axes>

```
[4]: # find range of 95th percentile
df2 = df[df["SBP"] >=df["SBP"].quantile(.95)]
max = df2["SBP"].max()
min = df2["SBP"].min()
print(f"95th percentile is SBP values in range of range ({min},{max})")
```

95th percentile is SBP values in range of range (164,230)