```
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import pandas as pd
from scipy import stats
```

Estimating the average weight of Women Olympians

```
df2 = pd.read_csv('athletes.csv')
df2.head()
```

	id	name	nationality	sex	dob	height	weight	sport	gold	silv
0	736041664	A Jesus Garcia	ESP	male	10/17/69	1.72	64.0	athletics	0	
1	532037425	A Lam Shin	KOR	female	9/23/86	1.68	56.0	fencing	0	
2	435962603	Aaron Brown	CAN	male	5/27/92	1.98	79.0	athletics	0	

```
df2 = df2[df2['sex'] == 'female']
df2.head()
```

	id	name	nationality	sex	dob	height	weight	sport	gold	s
1	532037425	A Lam Shin	KOR	female	9/23/86	1.68	56.0	fencing	0	
8	87689776	Aauri Lorena Bokesa	ESP	female	12/14/88	1.80	62.0	athletics	0	
9	997877719	Ababel Yeshaneh	ETH	female	7/22/91	1.65	54.0	athletics	0	

```
df2.shape
```

```
(5205, 11)
```

```
df2_samp = df2.sample(frac=0.1)    #draw the samples
df2_samp.shape
```

```
(520, 11)
```

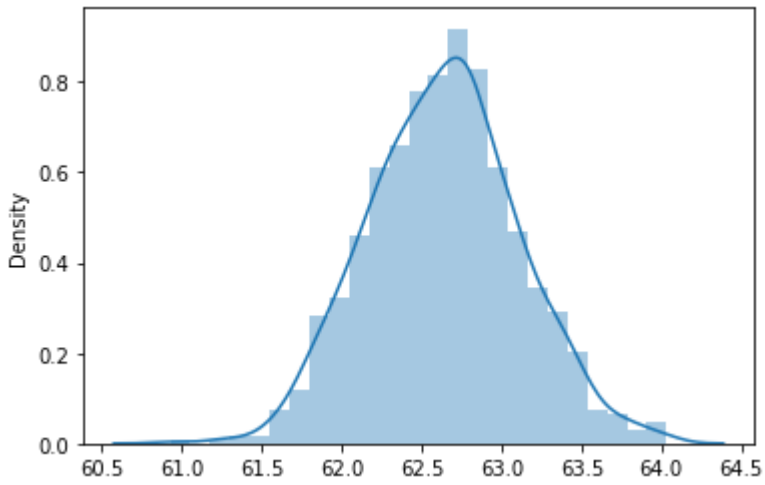
```
mean_hat = df2_samp['weight'].mean()     # sample mean
mean_hat
```

```
62.136546184738954
```

```
mean_hat_samples = []
for i in range(1000):
    sample = df2.sample(frac=0.1)
    mean_sample = sample['weight'].mean()
    mean_hat_samples.append(mean_sample)
```

```
sns.distplot(mean_hat_samples)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `di
  warnings.warn(msg, FutureWarning)
<matplotlib.axes._subplots.AxesSubplot at 0x7f5f8fa92190>
```




```
std_hat = df2_samp['weight'].std()/np.sqrt(len(df2_samp))  #sigma
std_hat
```

```
0.4644734919737783
```

```
z=stats.norm.ppf(q = 0.975)
z
```

```
1.959963984540054
```

```
lower_limit = mean_hat - z*std_hat
upper_limit = mean_hat + z*std_hat

lower_limit, upper_limit
```

```
(61.226194868696794, 63.046897500781114)
```

```
df2['weight'].mean()                    #interval
```

```
62.64442668797443
```

unknown SD

```
freedom=len(df2_samp) - 1
```

```
tscore=stats.t.ppf(0.975, freedom)
tscore
```

```
1.9645453314967687
```

```
stats.t.interval(alpha = 0.95,              # Confidence level
                 df= freedom,                    # Degrees of freedom
                 loc = mean_hat,         # Sample mean
                 scale = std_hat)
```

```
(61.22406695447786, 63.049025415000045)
```