

In [1]:

```
1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 import seaborn as sns
5 import scipy.stats as stats
6 from scipy.stats import sem, t
7 import statsmodels.api as sm
8 import os
```

In [2]:

```
1 df = pd.read_csv('C:/Users/MSKIT/Desktop/datasets/loans_income.csv')
2 df.head()
```

Out[2]:

	x
0	67000
1	52000
2	100000
3	78762
4	37041

In [3]:

```
1 loans_income = np.array(pd.read_csv('C:/Users/MSKIT/Desktop/datasets/loans_income.csv'))
2 loans_income[:5]
```

Out[3]:

```
array([[ 67000],
       [ 52000],
       [100000],
       [ 78762],
       [ 37041]], dtype=int64)
```

In [4]:

```
1 loans_income = np.array([item for sublist in loans_income for item in sublist])
```

In [6]:

```
1 def bootstrap(l,R):
2     n = len(loans_income)
3     means_of_boot_samples = []
4     for reps in range(R):
5         boot_sample = np.random.choice(loans_income, size = n)
6         means_of_boot_samples.append(round(np.mean(boot_sample),3))
7     return means_of_boot_samples
8 bootstrap(loans_income, 5)
9
```

Out[6]:

```
[68816.655, 68929.644, 68838.178, 69037.468, 68750.457]
```

In [7]:

```
1 np.std(bootstrap(loans_income, 100))
```

Out[7]:

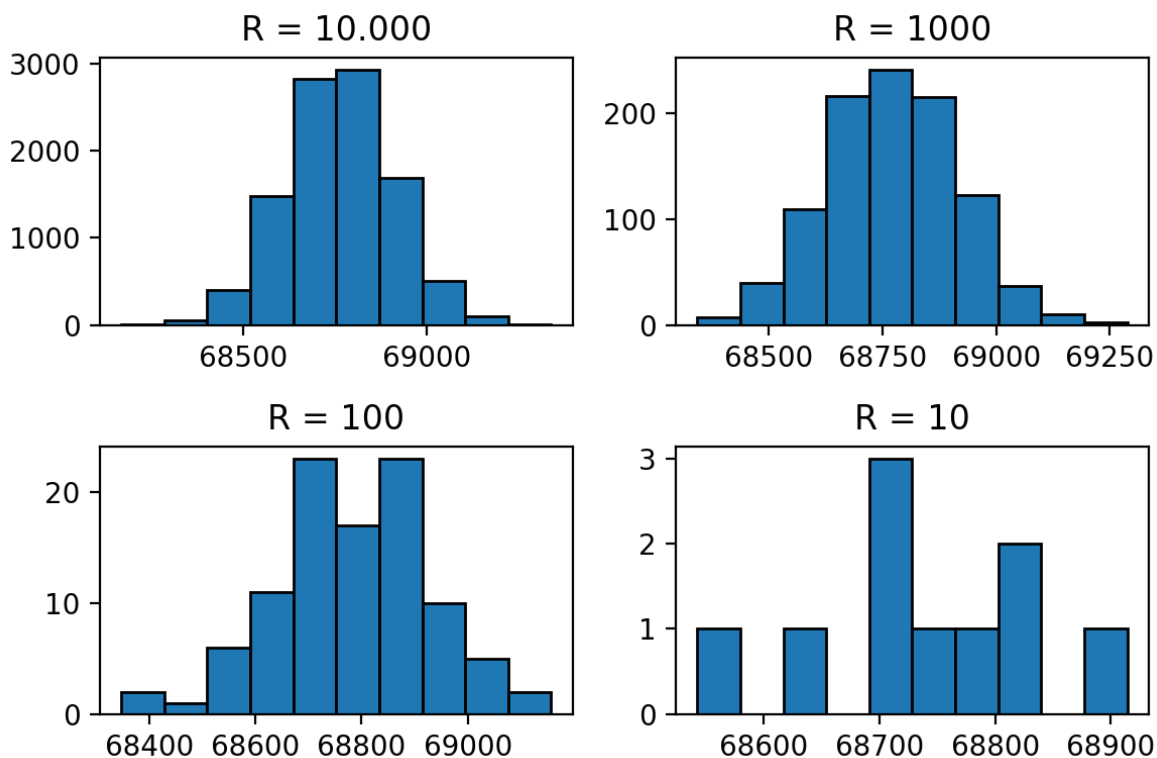
```
143.88873769343144
```

In [11]:

```

1 plt.figure(dpi= 200)
2 plt.subplot(221)
3 plt.title("R = 10.000")
4 plt.hist(bootstrap(loans_income,10000), edgecolor = 'k')
5
6 plt.subplot(222)
7 plt.title("R = 1000")
8 plt.hist(bootstrap(loans_income,1000), edgecolor = 'k')
9
10 plt.subplot(223)
11 plt.title("R = 100")
12 plt.hist(bootstrap(loans_income,100), edgecolor = 'k')
13
14 plt.subplot(224)
15 plt.title("R = 10")
16 plt.hist(bootstrap(loans_income,10), edgecolor = 'k')
17
18 plt.tight_layout()

```



In [12]:

```

1 data = bootstrap(loans_income, 1000)
2 lower_lim, upper_lim = np.percentile(data,2.5), np.percentile(data, 95)
3 print("Lower Limit: ", lower_lim)
4 print("Upper Limit: ", upper_lim)
5

```

Lower Limit: 68473.638375
Upper Limit: 69003.17105

In [14]:

```

1 plt.figure(dpi = 200)
2 plt.title("95% Confidence interval of loan applicants based on a sample of 1000 means")
3
4 sns.distplot(bootstrap(loans_income,1000), hist= True, kde=True,
5               color='darkblue', bins= 50,
6               hist_kws={'edgecolor':'black'},
7               kde_kws={'linewidth':2}
8               )
9 plt.axvline(x=lower_lim,color='red')
10 plt.axvline(x=upper_lim,color='red')

```

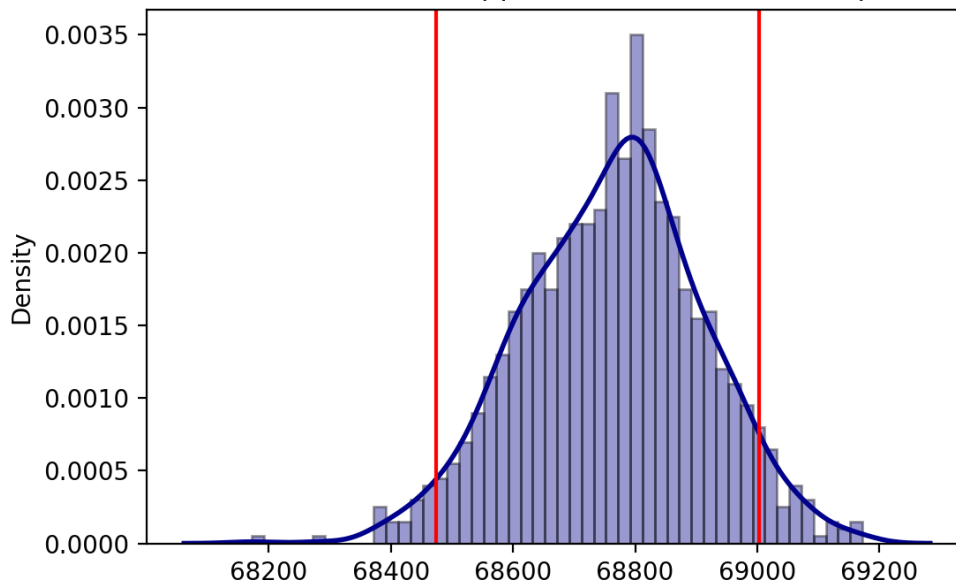
C:\Users\MSCIT\anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

warnings.warn(msg, FutureWarning)

Out[14]:

<matplotlib.lines.Line2D at 0x1afd9df6760>

95% Confidence interval of loan applicants based on a sample of 1000 means



In []:

1