In [1]:
```
import pandas as pd
import scipy.stats as stats
import matplotlib.pyplot as plt
import numpy as np

import math
```

In [3]:
```
df =pd.read_csv("class_heights.csv")
df
```

Out[3]:

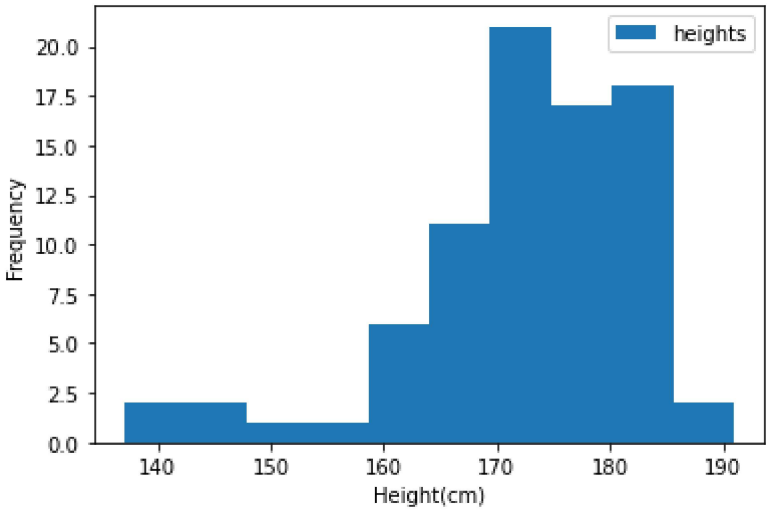
	gender	height
0	0	183
1	0	170
2	0	165
3	1	170
4	1	160
...	...	...
76	0	180
77	0	178
78	0	170
79	1	147
80	1	142

81 rows × 2 columns

In [4]:
```
students_heights = df.height
#print(students_heights)
```

In [5]:
```
plt.hist(students_heights,label='heights')

plt.xlabel('Height(cm)')
plt.ylabel('Frequency')
plt.legend()
```

Out[5]:
```
<matplotlib.legend.Legend at 0x20c3a5a0f70>
```



In [6]:
```
men = df.loc[df.gender==0, 'height']
#print(men)
```

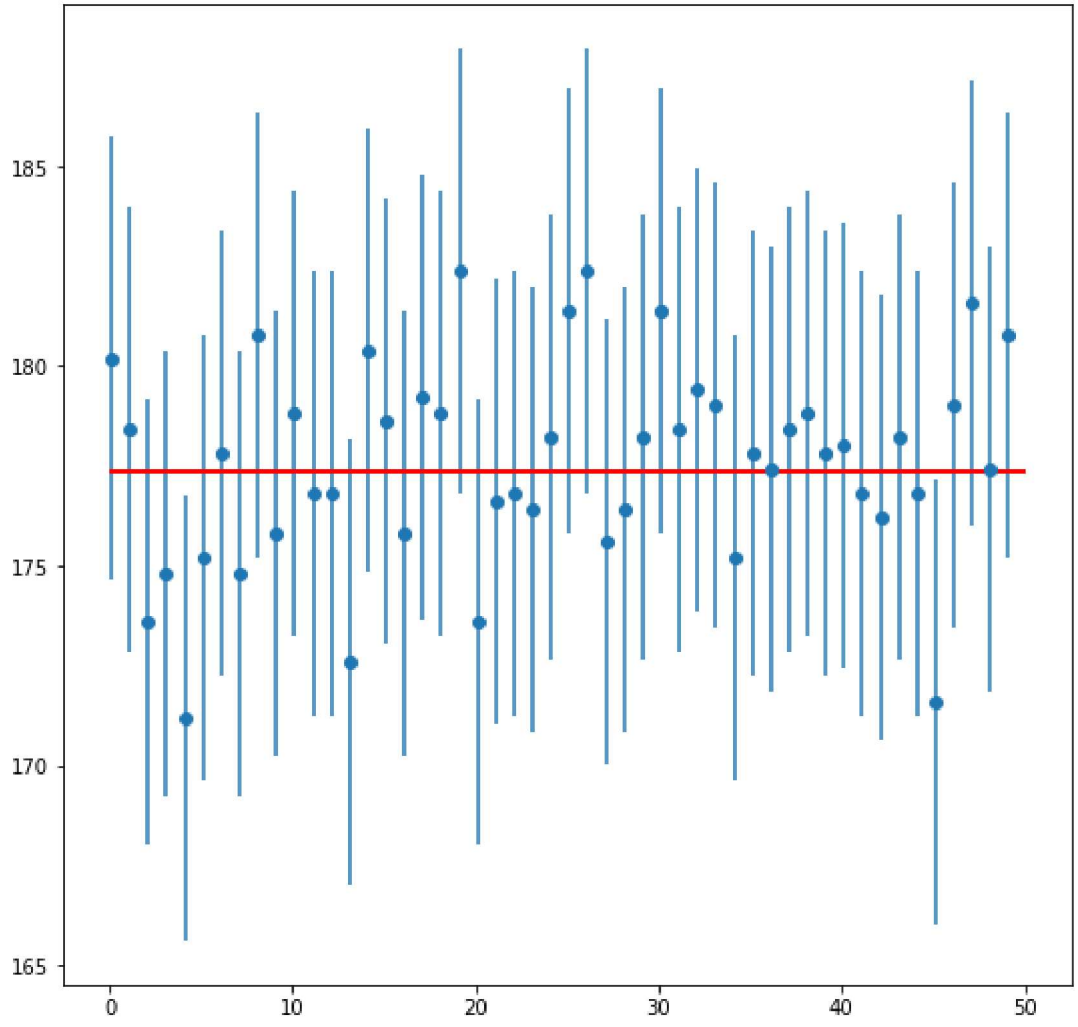
In [7]:
```
print(men.mean())
```

```
177.35087719298247
```

In [10]:
```
sample_size=5
intervals= []
sample_means =[]
for i in range(50):
    sample = np.random.choice(a=men,size=sample_size)
    #print(sample)
    sample_mean = sample.mean()
    #print(sample_mean)


    z_critical = stats.norm.ppf(q=0.975)
    #print(z_critical)

    men_std = men.std()
    #print(men_std)


    margin_of_error = z_critical * (men_std/math.sqrt(sample_size))
    #print(margin_of_error)


    confidence_interval = (sample_mean - margin_of_error, sample_mean + margin_of_error
    #print(confidence_interval)

    intervals.append(confidence_interval)
    sample_means.append(sample_mean)
```

In [11]:
```
plt.figure(figsize=(9,9))
plt.errorbar(x=np.arange(0.1,50,1),y=sample_means,yerr=[(top-bot)/2 for top,bot in inte


plt.hlines(xmin=0,xmax=50,y=177.35,linewidth=2.0,color='red')
```

Out[11]: <matplotlib.collections.LineCollection at 0x20c3a9572e0>



In [ ]: