MeanMedianMode

August 21, 2021

1 Mean, Median, Mode, and introducing NumPy

1.1 Mean vs. Median

Let's create some fake income data, centered around 27,000 with a normal distribution and standard deviation of 15,000, with 10,000 data points. (We'll discuss those terms more later, if you're not familiar with them.)

Then, compute the mean (average) - it should be close to 27,000:

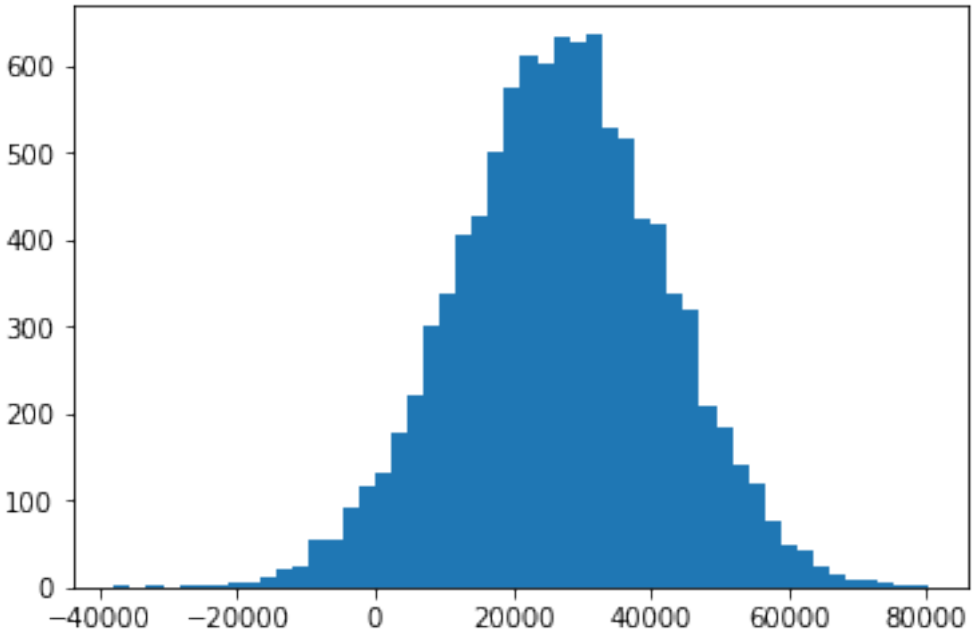
[1]:
```
import numpy as np

incomes = np.random.normal(27000, 15000, 10000)
np.mean(incomes)
```

[1]:
```
26933.720599154913
```

We can segment the income data into 50 buckets, and plot it as a histogram:

[2]:
```
%matplotlib inline
import matplotlib.pyplot as plt
plt.hist(incomes, 50)
plt.show()
```



Now compute the median - since we have a nice, even distribution it too should be close to 27,000:

[3]:
```
np.median(incomes)
```

[3]:
```
27092.39320043316
```

Now we'll add Jeff Bezos into the mix. Darn income inequality!

[4]:
```
incomes = np.append(incomes, [1000000000])
```

The median won't change much, but the mean does:

[5]:
```
np.median(incomes)
```

[5]:
```
27092.571036571815
```

[6]:
```
np.mean(incomes)
```

[6]:
```
126921.0284963053
```

1.2 Mode

Next, let's generate some fake age data for 500 people:

[9]:
```
ages = np.random.randint(18, high=90, size=500)
ages
```

[9]:
```
array([44, 51, 62, 80, 36, 61, 35, 64, 40, 45, 50, 33, 20, 63, 38, 87, 18,
       35, 21, 56, 57, 61, 52, 51, 45, 84, 40, 59, 22, 71, 25, 49, 19, 33,
       57, 43, 48, 65, 47, 38, 39, 59, 74, 49, 54, 59, 21, 79, 38, 80, 86,
       23, 26, 86, 55, 23, 34, 36, 48, 83, 77, 55, 22, 42, 78, 35, 53, 35,
       73, 87, 50, 85, 42, 52, 86, 31, 27, 22, 46, 20, 32, 63, 45, 61, 24,
       20, 45, 21, 40, 83, 66, 33, 49, 43, 78, 35, 83, 37, 48, 49, 81, 20,
       43, 26, 41, 24, 43, 69, 42, 39, 50, 57, 70, 31, 66, 82, 43, 19, 64,
       65, 29, 21, 35, 88, 43, 56, 56, 25, 19, 60, 57, 48, 74, 22, 65, 74,
       76, 42, 89, 59, 51, 19, 43, 80, 37, 20, 88, 30, 26, 37, 64, 28, 30,
       65, 21, 50, 54, 44, 86, 64, 81, 33, 67, 49, 55, 82, 64, 20, 47, 89,
       24, 18, 23, 32, 59, 86, 50, 25, 23, 23, 40, 41, 61, 38, 20, 20, 71,
       43, 30, 83, 83, 73, 49, 73, 20, 41, 84, 43, 39, 63, 25, 45, 33, 84,
       46, 70, 74, 57, 23, 37, 20, 78, 26, 25, 79, 38, 46, 53, 45, 27, 35,
       48, 83, 54, 50, 38, 23, 66, 55, 33, 62, 85, 36, 18, 22, 28, 22, 57,
       26, 62, 22, 23, 31, 35, 85, 48, 26, 41, 38, 25, 69, 44, 79, 26, 54,
       23, 19, 66, 41, 76, 60, 42, 39, 76, 27, 69, 57, 44, 86, 18, 42, 76,
       29, 86, 29, 56, 82, 65, 72, 68, 44, 83, 77, 37, 85, 34, 58, 86, 68,
       68, 36, 74, 78, 56, 19, 22, 60, 53, 30, 89, 41, 57, 60, 58, 84, 48,
       29, 85, 44, 25, 58, 69, 29, 80, 36, 34, 65, 58, 59, 66, 66, 42, 79,
       21, 72, 27, 54, 57, 55, 83, 37, 30, 63, 79, 56, 25, 35, 37, 31, 81,
       51, 63, 49, 31, 60, 77, 87, 20, 61, 49, 40, 25, 31, 87, 81, 33, 32,
       40, 88, 46, 54, 18, 62, 50, 44, 64, 83, 65, 47, 36, 45, 24, 36, 86,
       28, 57, 82, 45, 42, 61, 70, 35, 52, 21, 45, 81, 77, 52, 81, 48, 26,
       55, 69, 61, 84, 70, 71, 63, 79, 77, 25, 23, 43, 30, 21, 49, 31, 78,
       73, 53, 28, 48, 83, 31, 19, 62, 71, 20, 26, 23, 32, 77, 73, 58, 31,
       39, 20, 59, 39, 25, 51, 78, 62, 22, 48, 84, 75, 71, 24, 54, 68, 30,
       18, 30, 44, 51, 53, 25, 67, 39, 87, 73, 23, 67, 62, 83, 69, 59, 33,
       83, 70, 20, 51, 50, 36, 83, 48, 44, 89, 76, 64, 62, 65, 35, 73, 65,
       40, 75, 62, 20, 72, 35, 82, 35, 76, 72, 49, 55, 66, 88, 24, 72, 68,
       65, 79, 57, 76, 41, 68, 58])
```

[10]:
```
from scipy import stats
stats.mode(ages)
```

[10]:
```
ModeResult(mode=array([20]), count=array([15]))
```

[ ]:
```

```