

MeanMedianMode

December 20, 2016

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:

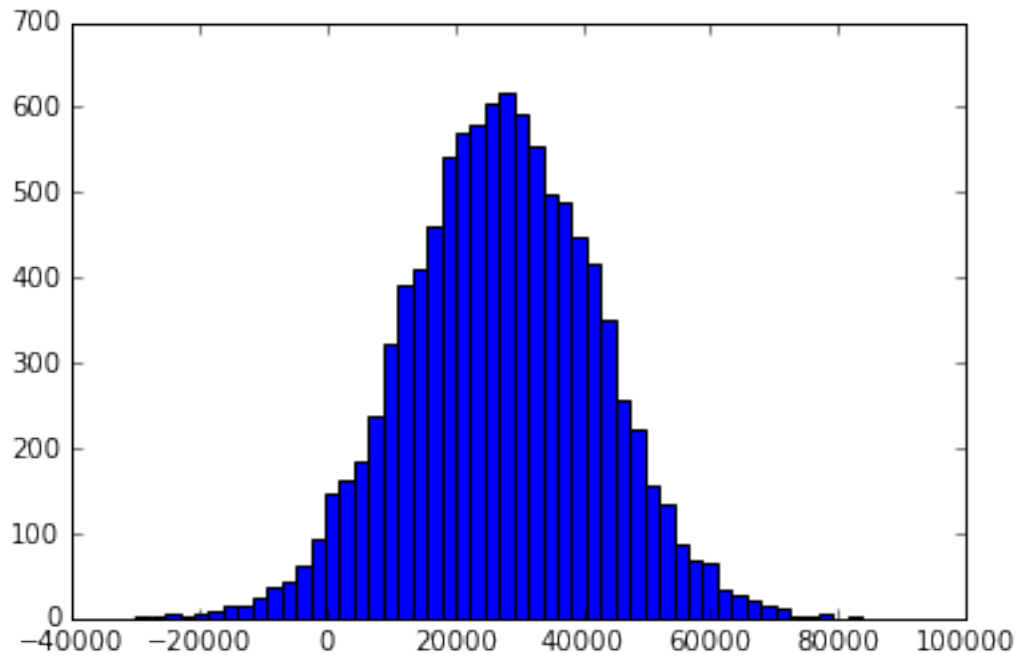
```
In [1]: import numpy as np
```

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

```
Out[1]: 27173.098561362742
```

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

```
In [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:

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

```
Out[3]: 27159.985229669175
```

Now we'll add Donald Trump into the mix. Darn income inequality!

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

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

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

```
Out[5]: 27163.131505581998
```

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

```
Out[6]: 127160.38252311043
```

1.2 Mode

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

```
In [7]: ages = np.random.randint(18, high=90, size=500)
ages
```

```
Out[7]: array([69, 87, 31, 22, 78, 37, 77, 32, 18, 59, 29, 43, 34, 33, 56, 83, 66,
              30, 77, 74, 31, 21, 85, 50, 47, 26, 72, 62, 33, 45, 86, 50, 86, 56,
              31, 84, 78, 27, 76, 42, 83, 64, 48, 54, 70, 56, 24, 50, 50, 71, 49,
              20, 85, 61, 33, 83, 55, 21, 60, 80, 56, 89, 61, 56, 52, 55, 20, 31,
              69, 50, 21, 52, 31, 83, 43, 77, 27, 67, 39, 39, 26, 38, 40, 73, 50,
              31, 87, 23, 50, 34, 69, 45, 83, 51, 88, 41, 64, 59, 40, 89, 57, 62,
              55, 75, 38, 51, 24, 21, 18, 75, 58, 62, 81, 65, 89, 64, 43, 33, 53,
              72, 20, 56, 19, 26, 81, 68, 70, 70, 41, 59, 50, 77, 62, 31, 87, 58,
              63, 83, 35, 55, 38, 85, 53, 66, 28, 74, 42, 28, 80, 69, 54, 25, 74,
              58, 27, 42, 87, 46, 43, 44, 33, 40, 21, 21, 73, 48, 87, 63, 84, 55,
              61, 66, 48, 73, 27, 60, 34, 77, 59, 58, 50, 70, 30, 76, 72, 33, 80,
              43, 63, 49, 60, 61, 53, 55, 79, 38, 46, 38, 81, 66, 29, 81, 46, 19,
              49, 57, 31, 18, 25, 47, 20, 88, 33, 88, 50, 22, 57, 39, 20, 59, 63,
              38, 35, 59, 28, 23, 56, 50, 46, 65, 46, 88, 87, 34, 73, 75, 32, 49,
              67, 77, 86, 38, 80, 36, 64, 79, 65, 51, 46, 54, 23, 82, 56, 41, 78,
              19, 45, 38, 70, 74, 56, 87, 49, 69, 30, 25, 22, 71, 39, 41, 46, 72,
              33, 72, 88, 37, 75, 39, 37, 21, 67, 86, 77, 20, 46, 53, 22, 85, 73,
              89, 67, 24, 24, 25, 62, 56, 58, 44, 63, 30, 36, 73, 49, 45, 26, 33,
              20, 62, 75, 34, 81, 59, 64, 27, 43, 23, 62, 75, 81, 40, 65, 29, 61,
              55, 81, 35, 68, 79, 86, 43, 35, 74, 59, 80, 75, 60, 82, 66, 54, 37,
              54, 71, 88, 46, 55, 63, 79, 89, 48, 61, 68, 78, 51, 32, 26, 48, 78,
              76, 62, 19, 19, 63, 20, 44, 28, 34, 58, 44, 36, 70, 34, 67, 50, 33,
              31, 18, 72, 55, 49, 63, 81, 65, 51, 46, 22, 55, 77, 76, 53, 79, 47,
              57, 46, 27, 29, 49, 71, 19, 85, 86, 77, 89, 59, 67, 26, 50, 79, 85,
              68, 51, 30, 18, 73, 52, 22, 53, 56, 26, 45, 60, 83, 50, 34, 68, 65,
              27, 72, 24, 34, 37, 52, 67, 79, 79, 24, 65, 71, 28, 29, 61, 34, 77,
              35, 59, 50, 83, 27, 32, 18, 81, 36, 46, 48, 39, 52, 23, 37, 62, 54,
              53, 50, 34, 36, 88, 83, 39, 89, 65, 83, 73, 66, 28, 36, 56, 86, 65,
              28, 46, 18, 61, 69, 80, 85, 29, 85, 44, 18, 61, 68, 83, 89, 53, 65,
              55, 66, 87, 55, 43, 32, 84])
```

```
In [8]: from scipy import stats
        stats.mode(ages)
```

```
Out[8]: ModeResult(mode=array([50]), count=array([16]))
```

```
In [ ]:
```