

Mean, Median, Mode, and introducing NumPy

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:

Mean Example

0, 2, 3, 2, 1, 0, 0, 2, 0

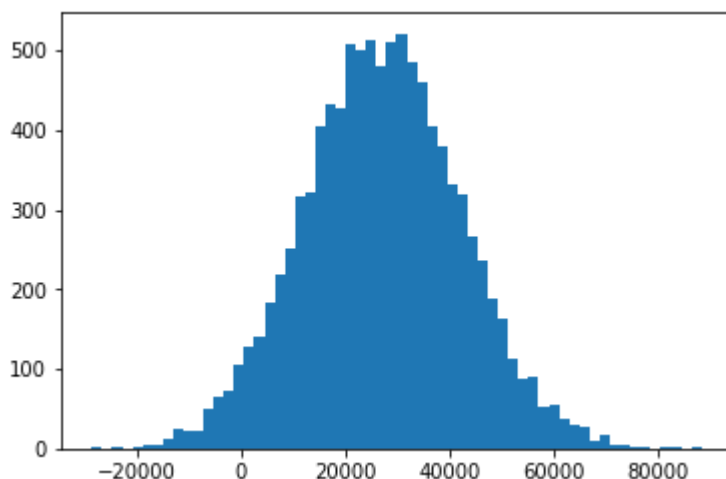
The MEAN is $(0+2+3+2+1+0+0+2+0) / 9 = 1.11$

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

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

Out[1]: 27065.854937711603

```
In [4]: import matplotlib.pyplot as plt
plt.hist(incomes, 60)
plt.show()
```



Median example

0, 2, 3, 2, 1, 0, 0, 2, 0

Sort it:

0, 0, 0, 0, 1, 2, 2, 2, 3



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

```
Out[5]: 26988.606055686752
```

Mode Example

0, 2, 3, 2, 1, 0, 0, 2, 0

How many of each value are there?

0: 4, 1: 1, 2: 3, 3: 1

The MODE is 0

generate some fake age data for 500 people

```
In [9]: from scipy import stats  
ages = np.random.randint(18, high=90, size=500)  
stats.mode(ages)
```

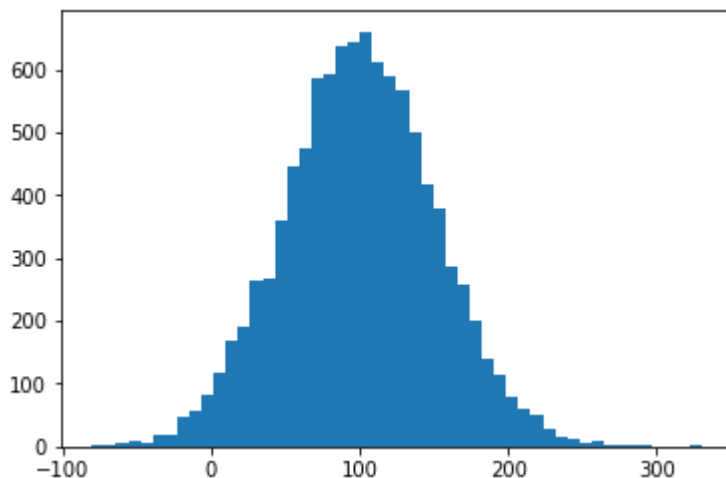
```
Out[9]: ModeResult(mode=array([31]), count=array([13]))
```

Standard Deviation and Variance

```
In [10]: import numpy as np
import matplotlib.pyplot as plt

incomes = np.random.normal(100.0, 50.0, 10000)

plt.hist(incomes, 50)
plt.show()
```



Variance

Variance (σ^2) is simply the **average of the squared differences from the mean**

Example: What is the variance of the data set (1, 4, 5, 4, 8)?

- First find the mean: $(1+4+5+4+8)/5 = 4.4$
- Now find the differences from the mean: $(-3.4, -0.4, 0.6, -0.4, 3.6)$
- Find the squared differences: $(11.56, 0.16, 0.36, 0.16, 12.96)$
- Find the average of the squared differences:
 - $\sigma^2 = (11.56 + 0.16 + 0.36 + 0.16 + 12.96) / 5 = 5.04$

```
In [11]: incomes.var()
```

```
Out[11]: 2554.924287394705
```

Standard Deviation (SD) is measured as the spread of data distribution in the given data set.

$$\text{Standard Deviation} = \sqrt{\text{mean}(\text{abs}(x - x.\text{mean}())^2)}$$

For example :

```
x = 1 1 1 1 1
Standard Deviation = 0 .

y = 9, 2, 5, 4, 12, 7, 8, 11, 9, 3, 7, 4, 12, 5, 4, 10, 9, 6, 9, 4
Step 1 : Mean of distribution 4 = 7
Step 2 : Summation of (x - x.mean())**2 = 178
Step 3 : Finding Mean = 178 /20 = 8.9
This Result is Variance.
Step 4 : Standard Deviation = sqrt(Variance) = sqrt(8.9) = 2.983..
```

In []: