Outline:

1. Writing short Python code using functions, loops, lists, numpy arrays, and dictionaries
2. Manipulating Python lists and numpy arrays and understanding the difference between them
3. Introducing the stats libraries scipy.stats and statsmodels

Getting Started with Python

Importing modules

All notebooks should begin with code that imports *modules*, collections of built-in, commonly-used Python functions. Below we import the Numpy module, a fast numerical programming library for scientific computing. Future labs will require additional modules, which we'll import with the same syntax.

`import MODULE_NAME as MODULE_NICKNAME`

In [28]:
```
import numpy as np #imports a fast numerical programming library
```

Now that Numpy has been imported, we can access some useful functions. For example, we can use `mean` to calculate the mean of a set of numbers.

In [29]:
```
my_list = [1.2, 2, 3.3]
np.mean(my_list)
```

Out[29]:
```
2.1666666666666665
```

Calculations and variables

In [30]:
```
# // is integer division
1/2, 1//2, 1.0/2, 3*3.2
```

Out[30]:
```
(0.5, 0, 0.5, 9.600000000000001)
```

The last line in a cell is returned as the output value, as above. For cells with multiple lines of results, we can display results using `print`, as can be seen below.

In [31]:
```
print(1 + 3.0, "\n", 9, 7)
5/3
```

```
4.0
 9 7
```

Out[31]:
```
1.6666666666666667
```

We can store integer or floating point values as variables. The other basic Python data types -- booleans, strings, lists -- can also be stored as variables.

In [32]:
```
a = 1
b = 2.0
```

Here is the storing of a list

In [33]:
```
a = [1, 2, 3]
```

Think of a variable as a label for a value, not a box in which you put the value

In [34]:
```
b = a
b
```

Out[34]:
```
[1, 2, 3]
```

This DOES NOT create a new copy of `a` . It merely puts a new label on the memory at a, as can be seen by the following code:

In [35]:
```
print("a", a)
print("b", b)
a[1] = 7
print("a after change", a)
print("b after change", b)
```

```
a [1, 2, 3]
b [1, 2, 3]
a after change [1, 7, 3]
b after change [1, 7, 3]
```

what if we use b=a.copy()?

Tuples

Multiple items on one line in the interface are returned as a *tuple*, an immutable sequence of Python objects. See the end of this notebook for an interesting use of `tuples` .

In [36]:
```
a = 1
b = 2.0
a + a, a - b, b * b, 10*a
```

Out[36]:
```
(2, -1.0, 4.0, 10)
```

`type()`

We can obtain the type of a variable, and use boolean comparisons to test these types. VERY USEFUL when things go wrong and you cannot understand why this method does not work on a specific variable!

In [37]:
```
type(a) == float
```
Out[37]:
```
False
```

In [38]:
```
type(a) == int
```
Out[38]:
```
True
```

In [39]:
```
type(a)
```
Out[39]:
```
int
```

EXERCISE 1: Create a tuple called `tup` with the following five objects:

- The first element is a float of your choice
- The second element is an integer of your choice
- The third element is the difference of the first two elements
- The fourth element is the sum of the first two elements
- The fifth element is the first element divided by the second element
- Display the output of `tup` .

Answer the following questions below:

- What is the type of the variable `tup` ?
- What happens if you try and chage an item in the tuple?

In [40]:
```python
# your code here
tup = 24.0, 4, 24.0-4, 24.0+4, 24.0/4
tup #This displays the output if you run the cell in jupyter as the last line in
print("tup:", tup) # This will print to console and output the contents in jupyte

# Running my own tests on setting variables and using them in list
t1 = 24.0
t2 = 4
tup2 = (t1, t2, t1-t2, t1+t2, t1/t2)
print("tup2:", tup2)

print("tup type:", type(tup))
print("contents in tup[0]", type(tup[0]))
print("contents in tup[1]", type(tup[1]))
print("contents in tup[2]", type(tup[2]))
print("contents in tup[3]", type(tup[3]))
print("contents in tup[4]", type(tup[4]))

# In analyzing the results I see that an operation between a float and integer
# is converted into a float result. ALso that a tuple is comma seperated on a sir
# line as stated earlier, but also can be held within two parenthesis.

#  1. tup is of the class tuple as shown below. Its a 5-tuple aka "quintuple"
# holding a float, integer, float, float, float.
# We know that tup is a tuple and the types of each piece
# through running the type function on tup and the tuple's elements
# A tuple is an immutable and ordered data type in Python.

tup2[1] = 100
print("Changed tup2[1] to 100:", tup2)

# 2. Trying to modify the contents in a previously instantiated tuple will raise
# a TypeError SUch that 'tuple' object does not support item assignment.
# In other words we may not reassign values in a tuple since its an immutable
# data type. We would have to mess around with converting to lists or other
# methods to work around such an issue.
```

```
tup: (24.0, 4, 20.0, 28.0, 6.0)
tup2: (24.0, 4, 20.0, 28.0, 6.0)
tup type: <class 'tuple'>
contents in tup[0] <class 'float'>
contents in tup[1] <class 'int'>
contents in tup[2] <class 'float'>
contents in tup[3] <class 'float'>
contents in tup[4] <class 'float'>
```

```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
<ipython-input-40-37c47492ed6b> in <module>
     24 # The tuple is an immutable and ordered data type in Python.
     25
---> 26 tup2[1] = 100
     27 print("Changed tup2[1] to 100:", tup2)
     28
```

```
TypeError: 'tuple' object does not support item assignment
```

Lists

Much of Python is based on the notion of a list. In Python, a list is a sequence of items separated by commas, all within square brackets. The items can be integers, floating points, or another type. Unlike in C arrays, items in a Python list can be different types, so Python lists are more versatile than traditional arrays in C or other languages.

Let's start out by creating a few lists.

In [41]:
```
empty_list = []
float_list = [1., 3., 5., 4., 2.]
int_list = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
mixed_list = [1, 2., 3, 4., 5]
print(empty_list)
print(int_list)
print(mixed_list, float_list)
```

```
[]
[1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
[1, 2.0, 3, 4.0, 5] [1.0, 3.0, 5.0, 4.0, 2.0]
```

Lists in Python are zero-indexed, as in C. The first entry of the list has index 0, the second has index 1, and so on.

In [42]:
```
print(int_list[0])
print(float_list[1])
```

```
1
3.0
```

What happens if we try to use an index that doesn't exist for that list? Python will complain!

In [43]:
```
# As stated above Python will throw an IndexError stating
# we are out of the range of the list
print(float_list[10])
```

```
---------------------------------------------------------------------------
IndexError                                Traceback (most recent call last)
<ipython-input-43-f9ddbd1267db> in <module>
      1 # As stated above Python will throw an IndexError stating we are out of
the range of the list
----> 2 print(float_list[10])

IndexError: list index out of range
```

You can find the length of a list using the built-in function `len` :

In [44]:
```
print(float_list)
len(float_list)
```

```
[1.0, 3.0, 5.0, 4.0, 2.0]
```

Out[44]:
```
5
```

Indexing on lists plus Slicing

And since Python is zero-indexed, the last element of `float_list` is

In [45]:
```
float_list[len(float_list)-1]
```

Out[45]:
```
2.0
```

It is more idiomatic in Python to use -1 for the last element, -2 for the second last, and so on

In [46]:
```
float_list[-1]
```

Out[46]:
```
2.0
```

We can use the `:` operator to access a subset of the list. This is called **slicing.**

In [47]:
```
print(float_list[1:5])
print(float_list[0:2])
```

```
[3.0, 5.0, 4.0, 2.0]
[1.0, 3.0]
```

In [48]:
```
lst = ['hi', 7, 'c', 'cat', 'hello', 8]
lst[:2]
```

Out[48]:
```
['hi', 7]
```

You can slice "backwards" as well:

In [49]:
```
float_list[:-2] # up to second last
# Note still prints forward, but does not include second last
```

Out[49]:
```
[1.0, 3.0, 5.0]
```

In [50]:
```
float_list[:4] # up to but not including 5th element
```

Out[50]:
```
[1.0, 3.0, 5.0, 4.0]
```

You can also slice with a stride:

In [51]:
```
float_list[:4:2] # above but skipping every second element
```

Out[51]:
```
[1.0, 5.0]
```

In [52]:
```
mylist = [1, 2, 3, 4, 5, 6, 7, 8, 9]
# This should print 1.0,5.0,2.0 since we should have whole array included
print(float_list[:5:2])
# The below will show that python does not care about the end limit even
# if its > len of array
print(float_list[:7:2])
# This demonstrates striding more clearly, it is hopping (::x) x-1 elements
# between each displayed element.
# In other words it is picking each x element after a displayed element.
print(float_list[:5:3])
print(float_list[:5:4])
print(mylist[:10:3])
```

```
[1.0, 5.0, 2.0]
[1.0, 5.0, 2.0]
[1.0, 4.0]
[1.0, 2.0]
[1, 4, 7]
```

We can iterate through a list using a loop. Here's a **for loop.**

In [53]:
```
for ele in float_list:
    print(ele)
```

```
1.0
3.0
5.0
4.0
2.0
```

What if you wanted the index as well?

In [54]:
```
# Then we could use enumerate and a for loop as shown below
```

Use the built-in python method `enumerate` , which can be used to create a list of tuples with each tuple of the form `(index, value)` .

In [55]:
```
for i, ele in enumerate(float_list):
    print(i, ele)
```

```
0 1.0
1 3.0
2 5.0
3 4.0
4 2.0
```

Appending and deleting

We can also append items to the end of the list using the `+` operator or with `append`.

In [56]:
```
float_list + [.333]
```

Out[56]:
```
[1.0, 3.0, 5.0, 4.0, 2.0, 0.333]
```

In [57]:
```
mylist + [10]
```

Out[57]:
```
[1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
```

In [58]:
```
print(mylist)
# Note that the + operator in single line when not stored to variable is not perm
```

```
[1, 2, 3, 4, 5, 6, 7, 8, 9]
```

In [59]:
```
#append permanently changes list
mylist.append(10)
print(mylist)
```

```
[1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
```

In [60]:
```
float_list.append(.444)
```

In [61]:
```
print(float_list)
len(float_list)
```

```
[1.0, 3.0, 5.0, 4.0, 2.0, 0.444]
```

Out[61]:
```
6
```

Now, run the cell with `float_list.append()` a second time. Then run the subsequent cell. What happens?

To remove an item from the list, use `del`.

In [62]:
```
# By running the previous two cells a second time we perform the append twice.
# The reults of this are: [1.0, 3.0, 5.0, 4.0, 2.0, 0.444, 0.444] with a length o
```

In [63]:
```
del(float_list[2])
print(float_list)
```

```
[1.0, 3.0, 4.0, 2.0, 0.444]
```

You may also add an element (elem) in a specific position (index) in the list

In [64]:
```
elem = '3.14'
index = 1
float_list.insert(index, elem)
float_list
```

Out[64]:
```
[1.0, '3.14', 3.0, 4.0, 2.0, 0.444]
```

In [65]:
```
# NOTE that inserting an element into a list will place that element at that posi
# and cause all subsequent elements to move up/back one index.
```

List Comprehensions

Lists can be constructed in a compact way using a *list comprehension*. Here's a simple example.

In [66]:
```
squaredlist = [i*i for i in int_list]
squaredlist
# THe above performs a square function for every element in the intlist [1, 2, 3,
```

Out[66]:
```
[1, 4, 9, 16, 25, 36, 49, 64, 81, 100]
```

In [67]:
```
And here's a more complicated one, requiring a conditional.
```

```
  File "<ipython-input-67-b5e13563f549>", line 1
    And here's a more complicated one, requiring a conditional.
        ^
SyntaxError: invalid syntax
```

In [68]:
```
comp_list1 = [2*i for i in squaredlist if i % 2 == 0]
print(comp_list1)
# From the above, this will double the value for any element in squared list that
```

```
[8, 32, 72, 128, 200]
```

This is entirely equivalent to creating `comp_list1` using a loop with a conditional, as below:

In [69]:
```
comp_list2 = []
for i in squaredlist:
    if i % 2 == 0:
        comp_list2.append(2*i)
print(comp_list2)
```

```
[8, 32, 72, 128, 200]
```

The list comprehension syntax

```
[expression for item in list if conditional]
```

is equivalent to the syntax

```
for item in list:
    if conditional:
        expression
```

Exercise 2: Build a list that contains every prime number between 1 and 100, in two different ways:

- 2.1 Using for loops and conditional if statements.
- 2.2 Using a list comprehension. You should be able to do this in one line of code. **Hint:** it might help to look up the function `all()` in the documentation.

In [70]:
```
### Your code here
# 2.1 For Loops and Conditional if Statements
# Note, must use list() in Python3 or above on range,
# and range is in format range(included, not included)
plist = []
hun_list = list(range(2,100))
print("Nearly Hundred List: "hun_list) # List from 2 to 99
# Note 1 and 100 are not prime so I excluded them with my approach
for dividend in hun_list:
    pflag = 1
    if dividend == 1:
        pflag = 0
    for divisor in range(2, dividend//2 + 1):
        if dividend % divisor == 0:
            pflag = 0
    if pflag:
        plist.append(dividend)
print("\nAll Primes Between 0 and 100")
print("\n 1.1 plist1:", plist)

# 2.2 Using list comp
plist2 = [i for i in hun_list if all(i % n != 0 for n in range(2, i))]
print("\n 1.2 plist2:", plist2)
```

```
[2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 2
3, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42,
43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62,
63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82,
83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99]
All Primes Between 0 and 100

 1.1 plist1: [2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 6
1, 67, 71, 73, 79, 83, 89, 97]

 1.2 plist2: [2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 6
1, 67, 71, 73, 79, 83, 89, 97]
```

Simple Functions

A *function* object is a reusable block of code that does a specific task. Functions are commonplace in Python, either on their own or as they belong to other objects. To invoke a function `func`, you call it as `func(arguments)`.

We've seen built-in Python functions and methods (details below). For example, `len()` and `print()` are built-in Python functions. And at the beginning, you called `np.mean()` to calculate the mean of three numbers, where `mean()` is a function in the numpy module and numpy was abbreviated as `np`. This syntax allows us to have multiple "mean" functions in different modules; calling this one as `np.mean()` guarantees that we will execute numpy's mean function, as opposed to a mean function from a different module.

User-defined functions

We'll now learn to write our own user-defined functions. Below is the syntax for defining a basic function with one input argument and one output. You can also define functions with no input or output arguments, or multiple input or output arguments.

```
def name_of_function(arg):
    ...
    return(output)
```

We can write functions with one input and one output argument. Here are two such functions.

In [71]:
```python
def square(x):
    x_sqr = x*x
    return(x_sqr)

def cube(x):
    x_cub = x*x*x
    return(x_cub)

square(5),cube(5)
```

Out[71]:
```
(25, 125)
```

What if you want to return two variables at a time? The usual way is to return a tuple:

In [72]:
```python
# Use a tuple as mentioned above.
def square_and_cube(x):
    x_cub = x*x*x
    x_sqr = x*x
    return(x_sqr, x_cub)

square_and_cube(5)
```

Out[72]:
```
(25, 125)
```

Lambda functions

Often we quickly define mathematical functions with a one-line function called a *lambda* function. Lambda functions are great because they enable us to write functions without having to name them, ie, they're *anonymous*. No return statement is needed.

In [73]:
```
# create an anonymous function and assign it to the variable square
square = lambda x: x*x
print(square(3))

hypotenuse = lambda x, y: x*x + y*y

## Same as
# def hypotenuse(x, y):
#     return(x*x + y*y)

hypotenuse(3,4)
```

```
9
```

Out[73]:
```
25
```

Methods

A function that belongs to an object is called a *method*. By "object," we mean an "instance" of a class (e.g., list, integer, or floating point variable).

For example, when we invoke `append()` on an existing list, `append()` is a method.

In other words, a *method* is a function on a specific *instance* of a class (i.e., *object*). In this example, our class is a list. `float_list` is an instance of a list (thus, an object), and the `append()` function is technically a *method* since it pertains to the specific instance `float_list`.

In [74]:
```
float_list = [1.0, 2.09, 4.0, 2.0, 0.444]
print(float_list)
float_list.append(56.7)
float_list
```

```
[1.0, 2.09, 4.0, 2.0, 0.444]
```

Out[74]:
```
[1.0, 2.09, 4.0, 2.0, 0.444, 56.7]
```

Exercise 3: generated a list of the prime numbers between 1 and 100

In Exercise 2, above, you wrote code that generated a list of the prime numbers between 1 and 100. Now, write a function called `isprime()` that takes in a positive integer N, and determines whether or not it is prime. Return `True` if it's prime and return `False` if it isn't. Then, using a list comprehension and `isprime()`, create a list `myprimes` that contains all the prime numbers less than 100.

In [75]:
```
newhunlist = list(range(2,100)) #create list from 2 to 99 , we know 1 and 100 are
def isprime(n):
    #Checks for all numbers from 2 to half of n or slightly past the half of n fo
    for i in range(2, n//2 + 1):
        if n % i == 0:
            #If it is divisible by something other then 1 and up to half of itsel
            return False
    # This is my else case, lets elminate the non prime numbers, else it is prime
    return True

myprimes = [i for i in newhunlist if isprime(i)]
print("Exercise 3: All primes < 100")
print("myprimes:", myprimes)
```

```
Exercise 3: All primes < 100
myprimes: [2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61,
67, 71, 73, 79, 83, 89, 97]
```

Introduction to Numpy

Scientific Python code uses a fast array structure, called the numpy array. Those who have programmed in Matlab will find this very natural. For reference, the numpy documention can be found here (https://docs.scipy.org/doc/numpy/reference/).

Let's make a numpy array:

In [76]:
```
my_array = np.array([1, 2, 3, 4])
my_array
```

Out[76]:
```
array([1, 2, 3, 4])
```

In [77]:
```
# works as it would with a standard list
len(my_array)
```

Out[77]:
```
4
```

The shape array of an array is very useful (we'll see more of it later when we talk about 2D arrays -- matrices -- and higher-dimensional arrays).

In [78]:
```
my_array.shape
```

Out[78]:
```
(4,)
```

Numpy arrays are **typed**. This means that by default, all the elements will be assumed to be of the same type (e.g., integer, float, String).

In [79]:
```
my_array.dtype
```

Out[79]:
```
dtype('int32')
```

Numpy arrays have similar functionality as lists! Below, we compute the length, slice the array, and iterate through it (one could identically perform the same with a list).

In [80]:
```
print(len(my_array))
print(my_array[2:4])
for ele in my_array:
    print(ele)
print(my_array[2:3]) # Testing the including of back part slicing again on array
```

```
4
[3 4]
1
2
3
4
[3]
```

There are two ways to manipulate numpy arrays a) by using the numpy module's methods (e.g., `np.mean()`) or b) by applying the function np.mean() with the numpy array as an argument.

In [81]:
```
print(my_array.mean())
print(np.mean(my_array))
```

```
2.5
2.5
```

A `constructor` is a general programming term that refers to the mechanism for creating a new object (e.g., list, array, String).

There are many other efficient ways to construct numpy arrays. Here are some commonly used numpy array constructors. Read more details in the numpy documentation.

In [82]:
```
np.ones(10) # generates 10 floating point ones
```

Out[82]:
```
array([1., 1., 1., 1., 1., 1., 1., 1., 1., 1.])
```

Numpy gains a lot of its efficiency from being typed. That is, all elements in the array have the same type, such as integer or floating point. The default type, as can be seen above, is a float. (Each float uses either 32 or 64 bits of memory, depending on if the code is running a 32-bit or 64-bit machine, respectively).

In [83]:
```
np.dtype(float).itemsize # in bytes (remember, 1 byte = 8 bits)
```

Out[83]:
```
8
```

In [84]:
```
np.dtype(int).itemsize #itemsize returns the bytes of data
```

Out[84]:
```
4
```

In [85]:
```
np.ones(10, dtype='int') # generates 10 integer ones
```

Out[85]:
```
array([1, 1, 1, 1, 1, 1, 1, 1, 1, 1])
```

In [86]:
```
np.zeros(10)
```

Out[86]:
```
array([0., 0., 0., 0., 0., 0., 0., 0., 0., 0.])
```

Often, you will want random numbers. Use the `random` constructor!

In [87]:
```
np.random.random(10) # uniform from [0,1]
```

Out[87]:
```
array([0.94759076, 0.47538248, 0.16562964, 0.89525144, 0.23710482,
       0.99106396, 0.54668684, 0.34559358, 0.51915205, 0.41494887])
```

You can generate random numbers from a normal distribution with mean 0 and variance 1:

In [88]:
```
normal_array = np.random.randn(1000)
print("The sample mean and standard devation are %f and %f, respectively." %(np.r
```

```
The sample mean and standard devation are -0.041333 and 1.015828, respectively.
```

In [89]:
```
len(normal_array)
```

Out[89]:
```
1000
```

You can sample with and without replacement from an array. Let's first construct a list with evenly-spaced values:

In [90]:
```
grid = np.arange(0., 1.01, 0.1)
#numpy.arange (start value, stop value not included, step size)
grid
```

Out[90]:
```
array([0. , 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1. ])
```

Without replacement

In [91]:
```
np.random.choice(grid, 5, replace=False)
```

Out[91]:
```
array([0. , 0.8, 0.4, 0.2, 0.3])
```

In [92]:
```
np.random.choice(grid, 5, replace=False)
```

Out[92]:
```
array([0.4, 0.6, 1. , 0.2, 0.9])
```

With replacement:

In [93]:
```
np.random.choice(grid, 20, replace=True)
```

Out[93]:
```
array([0. , 1. , 0.4, 0.5, 0.1, 0.7, 0.1, 0.3, 1. , 0.9, 0.4, 0.5, 0.4,
       0.2, 0.7, 0.7, 0.3, 0.5, 0.1, 0.9])
```

Numpy supports vector operations

What does this mean? It means that instead of adding two arrays, element by element, you can just say: add the two arrays.

In [94]:
```
first = np.ones(5)
second = np.ones(5)
first + second # adds in-place
```

Out[94]:
```
array([2., 2., 2., 2., 2.])
```

Note that this behavior is very different from python lists where concatenation happens.

In [95]:
```
first_list = [1., 1., 1., 1., 1.]
second_list = [1., 1., 1., 1., 1.]
first_list + second_list # concatenation
```

Out[95]:
```
[1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0]
```

On some computer chips, this numpy addition actually happens in parallel and can yield significant increases in speed. But even on regular chips, the advantage of greater readability is important.

Broadcasting

Numpy supports a concept known as *broadcasting*, which dictates how arrays of different sizes are combined together. There are too many rules to list here, but importantly, multiplying an array by a number multiplies each element by the number. Adding a number adds the number to each element.

In [96]:
```
first + 1
```

Out[96]:
```
array([2., 2., 2., 2., 2.])
```

In [97]:
```
first*5
```

Out[97]:
```
array([5., 5., 5., 5., 5.])
```

This means that if you wanted the distribution $N(5, 7)$ you could do:

In [98]:
```
normal_5_7 = 5 + 7*normal_array
np.mean(normal_5_7), np.std(normal_5_7)
```

Out[98]:
```
(4.7106693345556705, 7.110795762267876)
```

Multiplying two arrays multiplies them element-by-element

In [99]:
```
(first +1) * (first*5)
```

Out[99]:
```
array([10., 10., 10., 10., 10.])
```

You might have wanted to compute the dot product instead:

In [100]:
```
np.dot((first +1) , (first*5))
```
Out[100]:
```
50.0
```

#

Exercise 4: Matrix multiplication

Using numpy, create a random 5X5 matrix and multiply is by the 5X5 unit matrix

In [101]:
```
### Your code here
five_rand = np.random.randint(10, size=(5,5))
print("Random Matrix 5x5:\n",five_rand)
unit_mat = np.ones(shape=(5,5))
print("\nUnit Matrix 5x5:\n", unit_mat)
rand_unit_mult = five_rand * unit_mat
print("\nProduct of Random 5x5 Matrix and Unit Matrix:\n", rand_unit_mult)
```
```
Random Matrix 5x5:
 [[4 8 4 0 7]
 [7 4 0 9 5]
 [1 1 0 0 3]
 [1 8 1 6 1]
 [6 9 1 1 1]]

Unit Matrix 5x5:
 [[1. 1. 1. 1. 1.]
 [1. 1. 1. 1. 1.]
 [1. 1. 1. 1. 1.]
 [1. 1. 1. 1. 1.]
 [1. 1. 1. 1. 1.]]

Product of Random 5x5 Matrix and Unit Matrix:
 [[4. 8. 4. 0. 7.]
 [7. 4. 0. 9. 5.]
 [1. 1. 0. 0. 3.]
 [1. 8. 1. 6. 1.]
 [6. 9. 1. 1. 1.]]
```

Probabilitiy Distributions from `scipy.stats` and `statsmodels`

Two useful statistics libraries in python are `scipy` and `statsmodels` .

For example to load the z_test:

In [102]:
```
import statsmodels
from statsmodels.stats.proportion import proportions_ztest
```

In [103]:
```
x = np.array([74,100])
n = np.array([152,266])

zstat, pvalue = statsmodels.stats.proportion.proportions_ztest(x, n)
print("Two-sided z-test for proportions: \n","z =",zstat,", pvalue =",pvalue)
```

```
Two-sided z-test for proportions:
 z = 2.212695207500177 , pvalue = 0.026918666032452288
```

In [104]:
```
#The `%matplotlib inline` ensures that plots are rendered inline in the browser.
%matplotlib inline
import matplotlib.pyplot as plt
```

Let's get the normal distribution namespace from `scipy.stats`. See here for [Documentation (https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.norm.html)](https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.norm.html).

In [105]:
```
from scipy.stats import norm
```

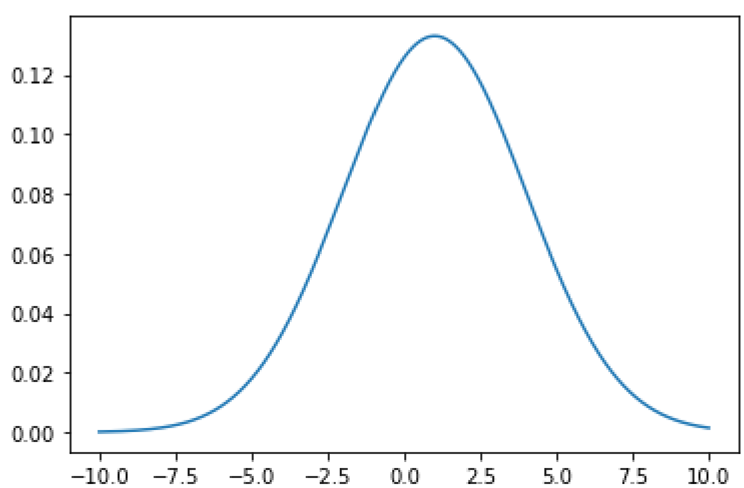
Let's create 1,000 points between -10 and 10

In [106]:
```
x = np.linspace(-10, 10, 1000) # linspace() returns evenly-spaced numbers over a
x[0:10], x[-10:]
```

Out[106]:
```
(array([-10.        , -9.97997998, -9.95995996, -9.93993994,
        -9.91991992, -9.8998999 , -9.87987988, -9.85985986,
        -9.83983984, -9.81981982]),
 array([ 9.81981982,  9.83983984,  9.85985986,  9.87987988,  9.8998999 ,
         9.91991992,  9.93993994,  9.95995996,  9.97997998, 10.        ]))
```

Let's get the pdf of a normal distribution with a mean of 1 and standard deviation 3, and plot it using the grid points computed before:

In [107]:
```
pdf_x = norm.pdf(x, 1, 3)
plt.plot(x, pdf_x);
```



And you can get random variables using the `rvs` function.

Referencies

A useful book by Jake Vanderplas: [PythonDataScienceHandbook (https://jakevdp.github.io/PythonDataScienceHandbook/)](https://jakevdp.github.io/PythonDataScienceHandbook/).

You may also benefit from using [Chris Albon's web site (https://chrisalbon.com)](https://chrisalbon.com) as a reference. It contains lots of useful information.

Dictionaries

A dictionary is another data structure (aka storage container) -- arguably the most powerful. Like a list, a dictionary is a sequence of items. Unlike a list, a dictionary is unordered and its items are accessed with keys and not integer positions.

Dictionaries are the closest data structure we have to a database.

Let's make a dictionary with a course number and their corresponding enrollment numbers.

In [108]:
```
enroll2021_dict = {'CS1': 500, 'CS2': 400, 'Stat1': 300, 'Stat2': 300, 'EE1': 40(
enroll2021_dict
```

Out[108]:
```
{'CS1': 500, 'CS2': 400, 'Stat1': 300, 'Stat2': 300, 'EE1': 400}
```

One can obtain the value corresponding to a key via:

In [109]:
```
enroll2021_dict['CS1']
```

Out[109]:
```
500
```

If you try to access a key that isn't present, your code will yield an error:

In [110]:
```
enroll2021_dict['CS148']
```

```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
<ipython-input-110-a5dad06f8832> in <module>
----> 1 enroll2021_dict['CS148']

KeyError: 'CS148'
```

Alternatively, the `.get()` function allows one to gracefully handle these situations by providing a default value if the key isn't found:

In [111]:
```
enroll2021_dict.get('CS148', 5)
```

Out[111]:
```
5
```

Note, this does not *store* a new value for the key; it only provides a value to return if the key isn't found.

In [112]:
```python
enroll2021_dict.get('CS148', None)
```

All sorts of iterations are supported:

In [113]:
```python
enroll2021_dict.values()
```

Out[113]:
```
dict_values([500, 400, 300, 300, 400])
```

In [114]:
```python
enroll2021_dict.items()
```

Out[114]:
```
dict_items([('CS1', 500), ('CS2', 400), ('Stat1', 300), ('Stat2', 300), ('EE1', 
400)])
```

We can iterate over the tuples obtained above:

In [115]:
```python
for key, value in enroll2021_dict.items():
    print("%s: %d" %(key, value))
```

```
CS1: 500
CS2: 400
Stat1: 300
Stat2: 300
EE1: 400
```

Simply iterating over a dictionary gives us the keys. This is useful when we want to do something with each item:

In [116]:
```python
second_dict={}
for key in enroll2021_dict:
    second_dict[key] = enroll2021_dict[key]
second_dict
```

Out[116]:
```
{'CS1': 500, 'CS2': 400, 'Stat1': 300, 'Stat2': 300, 'EE1': 400}
```

The above is an actual **copy** of *enroll2021_dict's* allocated memory, unlike, `second_dict = enroll2021_dict` which would have made both variables label the same memory location.

In the previous dictionary example, the keys were strings corresponding to course names. Keys don't have to be strings, though; they can be other *immutable* data type such as numbers or tuples (not lists, as lists are mutable).

Dictionary comprehension:

You can construct dictionaries using a *dictionary comprehension*, which is similar to a list comprehension. Notice the brackets {} and the use of `zip` (see next cell for more on `zip`)

In [117]:
```python
float_list = [1., 3., 5., 4., 2.]
int_list = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]

my_dict = {k:v for (k, v) in zip(int_list, float_list)}
my_dict
```

Out[117]:
```
{1: 1.0, 2: 3.0, 3: 5.0, 4: 4.0, 5: 2.0}
```

Creating tuples with `zip`

`zip` is a Python built-in function that returns an iterator that aggregates elements from each of the iterables. This is an iterator of tuples, where the i-th tuple contains the i-th element from each of the argument sequences or iterables. The iterator stops when the shortest input iterable is exhausted. The `set()` built-in function returns a `set` object, optionally with elements taken from another iterable. By using `set()` you can make `zip` printable. In the example below, the iterables are the two lists, `float_list` and `int_list`. We can have more than two iterables.

In [118]:
```python
float_list = [1., 3., 5., 4., 2.]
int_list = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]

viz_zip = set(zip(int_list, float_list))
viz_zip
```

Out[118]:
```
{(1, 1.0), (2, 3.0), (3, 5.0), (4, 4.0), (5, 2.0)}
```

In [119]:
```python
type(viz_zip)
```

Out[119]:
```
set
```

#

Exercise 5: Dictionary search

Given the dictionary `my_dict` that we made earlier, find all odd values, print the keys associated with this set of values, and then assign to the same key the value multiplied by 2.

In [120]:

```
## Your code here
print("keys: ", my_dict.keys())
print("\nvalues: ", my_dict.values())
print("\nitems: ", my_dict.items())
print("\n")
print("Keys Associated With Odd Values In my_dict Dictionary:")
for k, v in my_dict.items():
    # Check if the number is not divisible by 2, therefore odd
    if (v % 2 == 1):
        # Print the keys whose values were odd
        print(k)
        # Set the values for the odd value keys to twice their value, ie shift or
        my_dict[k] = float(int(v) << 1)
print("Modified Dictionary With Assigned Twice Value For Odd Values:\n", my_dict)
```

```
keys:  dict_keys([1, 2, 3, 4, 5])

values:  dict_values([1.0, 3.0, 5.0, 4.0, 2.0])

items:  dict_items([(1, 1.0), (2, 3.0), (3, 5.0), (4, 4.0), (5, 2.0)])


Keys Associated With Odd Values In my_dict Dictionary:
1
2
3
Modified Dictionary With Assigned Twice Value For Odd Values:
 {1: 2.0, 2: 6.0, 3: 10.0, 4: 4.0, 5: 2.0}
```

In [ ]:

In [ ]:

In [ ]: