Research Methods

Robert E Simpson

Singapore University of Technology & Design

robert_simpson@sutd.edu.sg

September 19, 2018

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Objectives

After the class you should be able to:

- Describe and plan algorithms using flow diagrams
- Define a function that takes an input, stores it as a local variable and *returns* an output to the main program
- Load and save `*.txt` or *.csv data to or from a python list
- Explain the difference between *local* and *global* variables
- Write a program that asks a user for an input
- Analyse distributions in python
- Write a routine to sequentially operate on each line of a list (CDFs)
- Plot x-y line plots, and bar charts

Local vs Global Variables

Global Variables

Global variables are accessible to any function in the program

Local Variables

Local variables only exist in a function whilst it is being run. After the function has finished, the local variables are not accessible to other parts of the program.

Looping through lines in a list

range `range([start], stop[, step])`
Generates a sequence of integers

- `start`: Starting number of the sequence.
- `stop`: Generate numbers up to, but not including this number.
- `step`: Difference between each number in the sequence.

x,y line plots `plt.plot(x_data,y_data)`

- First import the plotting library `import matplotlib.pyplot as plt`
- Add markers `plt.plot(my_x,pdf, marker="x")`
- Add x and y labels `plt.xlabel("x-name"` and `plt.ylabel("y-name"`

Bar Charts `plt.bar(x-coord, height)`

Looping through lines in a list

The following will loop through, and print each line of the list:

```
for line in my_list:
        print (line)
```

The print statement can be replaced with any operation that you want to perform on the lines of the list.

The following method can be used to write variable, strings, and lists to a list:

```
my_list=[]
for i in range(10):
        my_list.append(i)
```

Opening a file

The following will loop through, and print each line of the list:

```
with open('../Datasets/pingtime.txt','r') as f:
    output = f.read()
```

- In the first line, `'r'` tells python to open the file for read only.
- `f` is a file pointer.
- The full stop (`.`) tells the `read` function to operate on the file at which `f` points.

CSV reader

There is a useful python library for reading comma separated variable files. E.g.:

```
import csv
with open('file.csv','r') as f:
      reader = csv.reader(f, delimiter=',')
      mylist=list(reader)
```

The raw_input() command

```
person = raw_input('What is your name?: ')
age=raw_input('What is your age?: ')
print 'Hello', person,', you look much younger than',
    age,'years old!'
```

raw_input takes an input and stores it in a variable. In the above example, your name is stored in the variable "person" and your age is then stored in the variable "age".

Quick practice (optional)

recommended for beginners

a Write a function named `farenheit_to_celsius`. This function takes a Fahrenheit value as input and returns its centigrade equivalent. Recall: If C denotes the centigrade value and F its Fahrenheit equivalent, then $F = C \times 9/5 + 32$.

b $y(t) = v_0 t - \frac{1}{2} g t^2$ gives the height of a ball at time t when it is thrown with an initial velocity of v_0. Write a function named `position_velocity()` that takes v_0 and t as inputs and returns $y(t)$ and $y'(t)$, where $y'(t)$ is the first derivative of $y(t)$ with respect to t. Define the gravitational constant g, where appropriate? Round the output to two decimal number using `round(n,d)`. Use $g = 9.81\ m\ s^{-2}$. Round the output to two decimal number using `round(n,d)`. Use $g = 9.81\ m\ s^{-2}$

c Write a function that takes in a monthly saving amount and an annual interest rate, and returns the account value at the end of the sixth month. Write a test program that prompts the user to input his or her monthly saving amount and displays the account value. Round the final output to 2 decimal point.

Case problem 2.5

Means and Variances

Write functions to:

(a) Compute the mean of data stored in a list

(b) Compute the sample variance, population variance, and standard deviation of data stored in a list

Case problem 2.6

Major league baseball Player weights and heights

(a) Download `HeightsandWeight.txt` from the `Data Sets` folder on e-dimension and import it into python

(b) Use your previously written functions to compute the mean, sample standard deviation of the player weights and population standard deviation.

(c) Plot a histogram of the weights and heights. Do a sanity check (does the histrogram appear to have the same mean and standard deviation as that given by your function?)

The data was taken from HERE.

Case Problem 2.7

What is the chance of no rain for 20 days in Sept?

Records of Climate Station Mean

		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Rainfall	Mean Monthly Total (mm)	242.4	161.0	185.9	179.3	172.5	161.0	158.6	175.0	169.3	194.0	256.6	288.4
	Mean Raindays	15	11	14	15	15	13	13	14	14	16	19	19
Temperature °C	Mean Daily Maximum	30.1	31.1	31.6	31.7	31.6	31.3	30.9	30.9	30.9	31.1	30.6	30.0
	Mean Daily Minimum	23.3	23.6	23.9	24.4	24.8	24.8	24.6	24.5	24.2	24.1	23.8	23.5
	24-hr Mean	26.0	26.5	27.0	27.4	27.7	27.7	27.4	27.3	27.2	27.0	26.5	26.0

Case Problem 2.7

What is the chance of no rain for 20 days in Sept?

(a) Write a function to calculate the Binomial probability, $\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$

(b) Plot the probability distribution function for the number of rainy days in September.

(c) Plot the cumulative distribution function and use it to determine the probability that there are less than 12 dry days in September.

Case Problem 2.8:

(a) Write a function to computes the factorial of a number.

(b) Use your factorial function in a new function that calculates the Binomial Coefficient ("n choose x')

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

(c) Use these functions to plot the Binomial PDF for x heads in 5 coin tosses.

$$P(x, n, p) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

(d) Use these functions to plot the Binomial PDF for x heads in 50 coin tosses.