EASC2410 In-Class Exam 1: Python and Data Analysis Basics

2nd Semester
Department of Earth Sciences
Room 104, James Lee Science Building

Instructor: Dr. Binzheng Zhang

Student Name: _____ UID:_____

Exam Date: _____

Objective: Assess student understanding of basic knowledge of Python as a programing language, popular libraries for data analysis such as NumPy, Matplotlib and Pandas, and concepts/definitions in Data analysis and modeling

General Instructions: Read carefully each question; Again: Read carefully every item. The Ansering sheet is on **Page 2**.

- This is a **closed-book, written** exam;
- You have up to **75** minutes;
- Every item on the test awards **2** points for each correct answer, for a maximum possible score of **100** points;
- Non-graphing calculators are allowed;
- Smart phones, ipads, laptop computers are prohibited;
- Mere suspicion of cheating, discussions, using any unfair means of aid is enough to get your test withdrawn;
- Once a test is withdrawn, you get 0 points for the whole test;
- When you are done, turn in the exam including the problem sheet. Failure to do so will result in an automatic failing grade.

Anwering sheet

Student Name: _____ Score:_____

# of question	1	2	3	4	5	6	7	8	9	10
Answer										

# of question	11	12	13	14	15	16	17	18	19	20
Answer										

# of question	21	22	23	24	25	26	27	28	29	30
Answer										

# of question	31	32	33	34	35	36	37	38	39	40
Answer										

# of question	41	42	43	44	45	46	47	48	49	50
Answer										

1. Which of the follow is a part of Data Science?

   A. Data acquire B. Data prepare C. Data analyze D. All of the above

2. What can Python do in Data science?

   A. Data management B. Analytical processing
   C. Visualization D. All of the above

3. Which of the following is incorrect about comments?

   A. Comments start with "#"
   B. Comments are used to understand Python codes, functions, etc
   C. Python executes comments as regular codes
   D. Comments are ignored by Python

4. Which of the following is a reserved word in Python?

   A. del B. nan C. never D. All of the above

5. Which of the following is not a legal name for Python variables?

   A. Apples B. oRange_12 C. 2d_Melon D. my_durian

6. Which of the following is a floating point number in Python?

   A. 3 B. 3.0 C. '3.0' D. True

7. What is the output of the following Python code: `print(3e500)`

   A. 3e+500 B. 3.0e+500 C. 3e500 D. inf

8. What is the output of the following Python code: `print(True*2+False*True)`

   A. True B. False C. 2 D. 0

9. What is the output of the following Python code: `print(1.0+10%3)`

   A. 2.0 B. 2.3333333333 C. 4.0 D. 1.0

10. Which type of Python data structure is not ordered?

    A. list B. set C. tuple D. array

11. What is the output of the follwing Python code: `print('2.0'+3)`

    A. 5.0 B. '5.0' C. 5 D. TypeError

12. What is the output of the follwing Python code: `print('2.0'*3)`

    A. 6.0 B. '6.0' C. '2.02.02.0' D. TypeError

13. Which of the following code defines a Python list?

    A. x = [{'a', False, '400'}, ('spam', 42), [24 2]]

B. y = range(1,10)

C. z = 'Python' + "is" + "Fun"

D. None of the above

14. Define a set named "y":

```
y = {'a','2',400,'a',2,400}
```

what is the output for len(y)?

A. 6 B. 5 C. 4 D. sets have no length

15. What is the output of the following code?

```
z = ('a','2',400,'a',2,400)
z[2]=True
print(z)
```

A. ('a', '2', True, 'a', 2, 400)

B. ('a', True, 400, 'a', 2, 400)

C. ('2', True, 'a', 2, 400)

D. TypeError

16. What is the output of the following code?

```
myset1 = set(['Earth',9.8, 'Saturn', 10.3, 'Jupiter', 24.6])
# you can also define a set simply using curly brackets {}
myset2 = {'Mercury',5.4, 'Venus', 8.9, 'Earth', 9.8, 'Mars', 3.9}
print(myset1.intersection(myset2))
```

A. {'Earth', 9.8}

B. ['Earth', 9.8]

C. {'Saturn', 10.3, 'Jupiter', 24.6, 'Mercury',5.4, 'Venus', 8.9, 'Mars', 3.9}

D. ['Saturn', 10.3, 'Jupiter', 24.6, 'Mercury',5.4, 'Venus', 8.9, 'Mars', 3.9]

17. How to access the word "Earth Sciences" in the following string variable b?

```
b = "HKU Earth Sciences"
```

A. b[5:] B. b[4:18] C. b(4:) D. b[-4:]

18. What is the output of the following code?

```
print(('Earth'=='Mars')or(6380>=3350))
```

A. True B. False C. 6380 D. 3350

19. What is the output after calling the following if-elif-else statement?

```
mylist=['jane','josh','sid','geoff'] # names

if 'josh' in mylist: # check josh
    print ('josh in list!')
elif 'george' in mylist: # check george
    print ('george in list')
else: # see if both are in list
    print ("josh and george isn't in list")
```

A. 'josh in list!'

B. 'george in list'
C. 'josh in list!'
'george in list'
D. 'josh in list!'
'george in list'
'josh and george isn't in list'

20. What is the output after calling the following while statement?

```
mylist=['jane','josh','sid','geoff'] # names

while 'susie' not in mylist: # check if 'susie' is NOT in the list
    print(mylist)
    mylist.append('susie')              # append 'susie' to mylist
```

A. ['jane', 'josh', 'sid', 'geoff']
B. ['jane', 'josh', 'sid', 'geoff', 'susie']
C. No outputs
D. Indentation Error

21. What is the output after calling the following nested loops:

```
solar_planets = {'Mercury':440.0, 'Venus' :737.1,
                 'Earth'  :288.5, 'Mars'  :210.2,
                 'Jupiter':110.0, 'Saturn':81.2}

habitable = False

for key in solar_planets.keys():

    temperature = solar_planets[key]

    if (temperature <= 300) and (temperature >= 278):
        habitable = True
    else:
        habitable = 3.0

print(habitable)
```

A. False B. True C. 3.0 D. None of the above

22. What is the output after calling the following function convertF2C( ):

```
def convertF2C(in_args = 95):
    """
    This code convert Fahrenheit (F) to Celsius (C)
    INPUT    : temperature in F
    OUTPUT   : temperature in C
    Algorithm: C = 9/5*(F-32)
    """
    out_args = (in_args - 32.0)*5.0/9.0
    return out_args

# now lets call the function here by simply type in the function name
convertF2C()
```

A. NameError - name 'in_args' is not defined
B. 95
C. 35.0
D. Syntax Error

23. What is the correct syntax for importing the "numpy" module and name it as "pn"?

A. import numpy as np
B. from numpy import pn
C. import numpy as pn
D. from numpy import np as pn

24. Which of the following is an invalid Python statement?

A. abc = 1,000,000
B. a b c = 1000 2000 3000
C. a, b, c = 1000, 2000, 3000
D. a_b_c = 1000, 2000, 3000

25. What is the output from the following code?

```
height = [1,3,5,7,9]
print(height**2)
```

A. [2, 6, 10, 14, 18]
B. [1, 9, 25, 49, 81]
C. [1, 3, 5, 7, 9, 1, 3, 5, 7, 9]
D. TypeError

26. Which of the following function is not a method of a NumPy object?

A. min( )
B. median( )
C. std( )
D. scatter( )

27. What is the output of the following Python code?

```
from numpy import *

height = array([1.73, 1.68, 1.71, 1.89, 1.79])
weight = array([63.1, 55.6, 65.2, 89.8, 78.5])

tall = height > 1.79
weight[tall]
```

A. array([False, False, False, True, False])
B. array([1.89])
C. array([89.8])
D. array([True])

28. What is the length of the 1-D array num: num = np.arange(1,10,2)?

A. 4
B. 5
C. 9
D. 10

29. In the following 2-D NumPy array x, how to access the sub-array denoted by the red box?

```
Out[84]: array([[0.49, 0.02, 0.68, 0.81, 0.37, 0.53, 0.62],
                [0.73, 0.43, 0.07, 0.53, 0.51, 0.36, 0.28],
                [0.3 , 0.46, 0.36, 0.8 , 0.15, 0.33, 0.93],
                [0.22, 0.12, 0.54, 0.59, 0.17, 0.03, 0.22],
                [0.21, 0.21, 0.67, 0.19, 0.28, 0.22, 0.96],
                [0.24, 0.79, 0.67, 0.53, 0.42, 0.76, 0.29],
                [0.39, 0.47, 0.42, 0.22, 0.77, 0.82, 0.72],
                [0.06, 0.63, 0.45, 0.23, 1.  , 0.39, 0.84]])
```

A. x[4:6, 2:4]
B. x[5:7, 2:4]
C. x[4:6, 3:5]
D. x[5:7, 3:5]

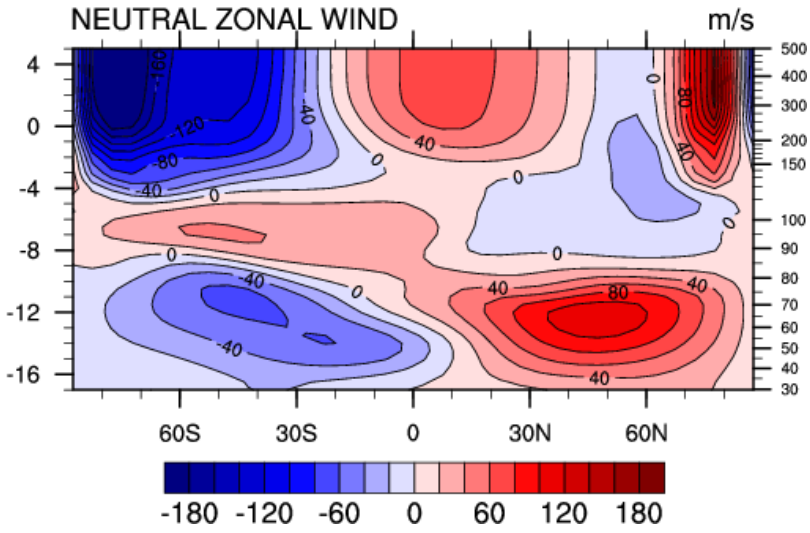
30. When using the plot( ) function from the matplotlib.pyplot module, which of the following function is used to change the location of the ticks and corresponding labels of the ticks?

A. matplotlib.pyplot.xlabel( )
B. matplotlib.pyplot.xlim( )
C. matplotlib.pyplot.xtick( )
D. matplotlib.pyplot.legend( )

31. What is the general name of the following three colormaps?

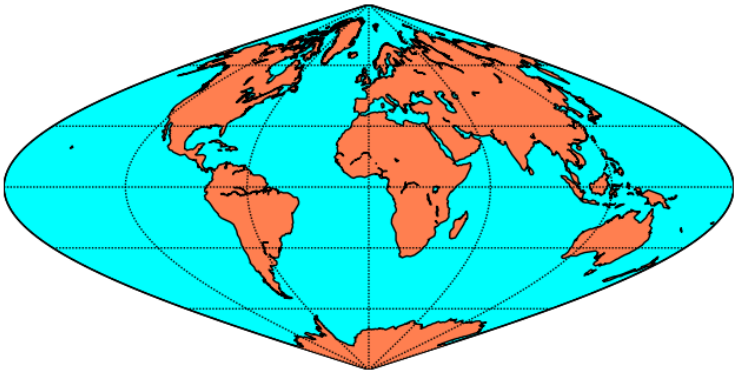
A. sequential colormaps
B. diverging colormaps
C. cyclic colormaps
D. qualitative colormaps

32. What matplotlib function is used to generate the following 2-D color plot?



A. pcolor( )
B. contour( )
C. contourf( )
D. plot_surface( )

33. What is the name of the following map projection from Basemap?



A. Mercator Projection
B. Orthographic Projection
C. Lambert Conformal Projection
D. Sinusoidal Projection

34. Which of the following description is incorrect about the following NumPy command?

```
EQ = np.loadtxt('datasets/earthquake.csv', delimiter=",")
```

A. "earthquake.csv" is the name of a data file
B. "datasets" is the name of a folder
C. "datasets/earthquake.csv" is an absolute path
D. delimiter="," means the data are seperated by a string ","

35. Given the following DataFrame named "brics", which of the following description is incorrect?

	Country	Population	Area	Capital
BR	Brazil	200	8515767	Brazilia
RU	Russia	144	17098242	Moscow
IN	India	1252	3287590	New Delhi
CH	China	1357	9596961	Beijing
SA	South Africa	55	1221037	Pretoria

A. "brics" is a 2-D Data Frame
B. the column indices are ["Country", "Population", "Area", "Captial"]
C. the row indices are integers [0, 1, 2, 3, 4]
D. brics.Population gives the population of all the countries

36. In the above data frame "brics", how to access the Capital city of India?

A. brics.Capital['IN']
B. brics.loc['IN']['Capital']
C. brics['Capital']['IN']
D. brics['IN']['Capital']

37. Which of the following Pandas function can be used to bin data in a data frame?

A. head( )
B. describe( )
C. cut( )
D. corr( )

38. Which of the following number is a precise but inaccurate approximation of π?

A. 3.1415926
B. 3.1
C. 3.58936270
D. 4

39. Suppose you play a fair coin flipping game with a friend following these rules:

- if you get "Head", you win $10
- if you get "Tail", you lose $5

What is your mathematical expectation of winning $ from this game?

A. $ 0
B. $ 2.5
C. $ -2.5
D. $ 5

40. For probability distributions, which of the following statement is incorrect?

A. Probability distributions are lists of outcomes and their associated probabilities
B. A function that represents a continuous probability distribution is called a probability mass function
C. The values in a probability mass function represent actual probabilities
D. The area under the curve produced by a probability density function represents a probability

41. What is the name of the theoretical probability distribution defined by following lambda function?

```
from scipy.special import factorial

Probability = lambda x,n,p :(factorial(n)/(factorial(x)*factorial(n-x)))*(p**(x))*(1.-p)**(n-x)
```

A. Binomial
B. Uniform
C. Normal
D. Empirical

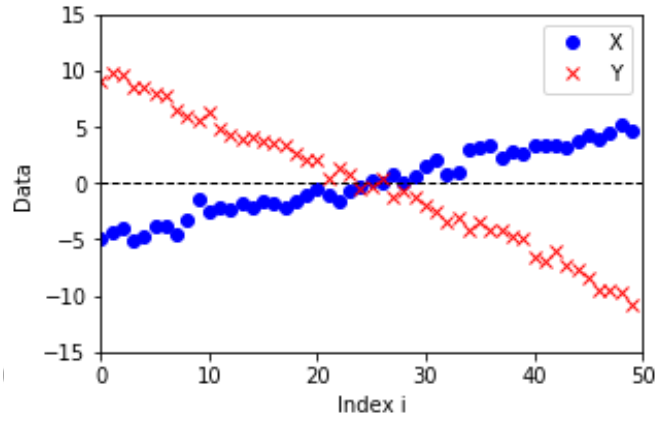
42. Which of the following descriptions about a Normal distribution is incorrect?

A. A Normal distributions is symmetric about its mean value μ
B. The Standard deviation σ shows the spread of the Normal distribution
C. The probability of a Normal distirbution within μ±σ is approximately 68%
D. The mean and median of a Normal distribution are not the same

43. In a positively skewed distribution, what is the relationship between mean, median and mode?

A. Mean = Median = Mode
B. Mean > Median > Mode
C. Mean < Median < Mode
D. Mean > Median = Mode

44. Which description of the following two datasets X and Y is incorrect?



A. X and Y are linearly dependent
B. the covariance between X and Y is negative
C. the Pearson correlation coefficient between X and Y is close to -1
D. X and Y are not correlated

45. Which descripition of time series is incorrect?

A. Time series data is also known as longitudinal data
B. Time series are usually defined in the time domain
C. Some time series data sets have no trend
D. They may have seasonalility, cyclic and/or random variations

46. Given the following time series y(t) defined in 0 < t < 10 (t in seconds), which of the following property is correct?

$$y(t) = 1.0 \sin(4\pi t) + 0.3 \sin(6\pi t)$$

A. y(t) has two frequency components with 2π Hz and 3π Hz
B. the leading frequency component in y(t) has a power of 0.5
C. y(t) is a periodic signal with a period of 12π
D. y(t) is not a periodic signal

47. In the following code defining a time series y(t) with 0 < t < 4 (t in second), what is the sampling frequency f_s of the time series y(t)?

```
n = np.array(range(1,8))

freq = n
magn = 1/n

t = np.arange(0,4,0.05)

y = t*0
for n in range(len(freq)):
    y = y+magn[n]*np.sin(2*np.pi*freq[n]*t)
```

A. 0.05 B. 1.0 C. 4.0 D. 20

48. A periodogram is:

A. a time series
B. a periodic spectrum
C. a power spectrum
D. a statistical distribution

49. In a time series, which of the following component is usually predictable?

A. trend B. seasonality C. cyclic D. irregular

50. In a monthly mean temperature dataset (i.e., one data point recorded for each month, or $\Delta t = 1$ month), in order to remove seasonal variations using the rolling( ) function from Pandas, what is an appropriate size of the rolling window?

A. 3 B. 6 C. 12 D. 24