NATIONAL UNIVERSITY OF SINGAPORE

DAO2702 PROGRAMMING FOR BUSINESS ANALYTICS

(Sample Exam)

Time Allowed: **2 Hours**

INSTRUCTIONS TO STUDENTS

1. Please write your student number only. **Do not write your name**.
2. This assessment paper contains **21** multiple-choice questions and **1** quantitative question, totally **12** printed pages.
3. Students are required to answer **ALL** questions.
4. This is a CLOSED BOOK examination (with authorised materials).
5. Specific permitted materials is **one (1)** A4-sized help sheet.
6. Write your answers in the boxes provided after each question (or part of a question). *You should plan your answers to ensure that they fit within the spaces provided. Other than this cover page and the spaces designated for providing your answers, you may do your "rough work" anywhere. Whatever you write outside of the answer boxes may be ignored.*

Student No :

Question	Max	Mark
PART I	63	
PART II	37	
Total	100	

PART I. MULTIPLE-CHOICE QUESTIONS (63 marks)

There is only **one** correct answer in each question. Please shade your answer in the bubble card given. **Each question accounts for 3 marks.**

1. What is the output message of the following code?

```
a = '2.5'
b = "1"

print(str(a*2) + str(b))
```

A. 2.52.51
B. 6
C. 51
D. An error message

2. What is the output message of the following code?

```
a = 3
b = 4

statement1 = a <= 3 and a > 2
statement2 = b > 2 and b <= 3
print(statement1 != statement2)
```

A. True
B. False
C. 3
D. 4

3. What is the printed message of the following code?

```
letters = 'AaBbCcDdEeFfGg'

new_str=''
for letter in letters:
    if letter in 'ABCD':
        continue
    new_str = new_str + letter
    if letter in 'efg':
        break

print(new_str)
```

A. abcde
B. abcDdEe
C. abcdEe
D. abcdE

4. What is the printed message of the following code?

```
name = 'John Fitzgerald Kennedy'

while True:
    index = name.find(' ')
    if index == -1:
        break
    name = name[index+1:]

print(name)
```

A. Kennedy
B. Fitzgerald Kennedy
C. JFK
D. An error message

5. What is the output message of the following code?

```
list_1 = [1, 2]
list_2 = list_1
list_1.insert(1, '0')
list_2 += list_1

print(list_2)
```

A. ['0', 1, 2, '0', 1, 2, '0', 1, 2]
B. [1, '0', 2, 1, '0', 2, 1, '0', 2]
C. [1, '0', 2, 1, 2, 1, 2]
D. [1, '0', 2, 1, '0', 2]

6. What is the output message of the following code?

```
sentence = "Business analytics"

print(sentence[-8:-3])
```

A. analyt
B. nalyt
C. analyti
D. nalyti

7. What is the output message of the following code?

```
list1 = [1, 2, 3]
list2, list3 = list1, list1[1:]
list2.pop()

print(list2+list3)
```

A. [1, 2, 2]
B. [1, 2, 2, 3]
C. [1, 2, 3, 2, 3]
D. [1, 2, 3, 1, 2, 3]

8. What is the output message of the following code?

```
list_1 = [True, "2.0", 3.5]
list_2 = list_1 * 2
list_1.remove(3.5)
list_2.remove("2.0")
print(list_1 + list_2)
```

A. [True, '2.0', True, 3.5, True, '2.0', 3.5]
B. [True, True, True]
C. [True, '2.0', True, 3.5, True, 3.5]
D. [True, '2.0', True, True]

9. The code below

```
a, b = 2.5, 3.9

a = a + b
b = a - b
a = a - b
```

is equivalent to which of the following code?

A.
```
a, b = 2.5, 3.9

a, b = b, a
```

B.
```
a, b = 2.5, 3.9

a, b = a + b, a - b
```

C.
```
a, b = 2.5, 3.9

a, b = a - b, a – b
```

D.
```
a, b = 2.5, 3.9

b, a = a, a - b
```

10. Given a list “var” of variance measures, which of the following code is correct in transforming the variances to standard deviations?

A.
```
var = [3.5, 6.2, 7.3, 8.5]
std = var ** 0.5
```

B.
```
var = [3.5, 6.2, 7.3, 8.5]
std = []
for item in var:
    std = std + item ** 0.5
```

C.
```
var = [3.5, 6.2, 7.3, 8.5]
std = []
for item in var:
    std.append([item ** 0.5])
```

D.
```
import numpy as np

var = [3.5, 6.2, 7.3, 8.5]
std = np.array(var) ** 0.5
std = list(std)
```

11. What is the output message of the following code?

```
def sum_two(x, y, z):
    """This is a function"""

    return x+y, y+z, x+z

"""Outside the function"""
x = 1
y = 2
z = 3
result = sum_two(x, y, z)
a, b = result[::2]
print(str(a) + str(b))
```

A. 34
B. 35
C. 45
D. An error message

12. The function “pie” from “matplotlib.pyplot” is used to draw pie graphs.

```
import matplotlib.pyplot as plt
help(plt.pie)
```

```
Help on function pie in module matplotlib.pyplot:

pie(x, explode=None, labels=None, colors=None, autopct=None, pctdistance=0
.6, shadow=False, labeldistance=1.1, startangle=None, radius=None, counter
clock=True, wedgeprops=None, textprops=None, center=(0, 0), frame=False, r
otatelabels=False, *, data=None)

……
Parameters
    ----------
    x : array-like
        The wedge sizes.
```

```
    explode : array-like, optional, default: None
        If not *None*, is a ``len(x)`` array which specifies the fraction
        of the radius with which to offset each wedge.

    labels : list, optional, default: None
        A sequence of strings providing the labels for each wedge

    colors : array-like, optional, default: None
        A sequence of matplotlib color args through which the pie chart
        will cycle.  If *None*, will use the colors in the currently
        active cycle.
```

.....

Which of the following code is correct in drawing a pie without shadows, where the wedge sizes are specified by the list "grades", the labels are specified by the list "labels", and the explode distance is `[0.1]*7`

A.
```
grades = [0.02, 0.08, 0.15, 0.25, 0.2, 0.15, 0.15]
labels = ['A+', 'A', 'A-', 'B+', 'B', 'B-', 'C']

plt.pie(x=grades, [0.1]*7, shadow=False, labels=labels)
plt.show()
```

B.
```
grades = [0.02, 0.08, 0.15, 0.25, 0.2, 0.15, 0.15]
labels = ['A+', 'A', 'A-', 'B+', 'B', 'B-', 'C']

plt.pie(labels=labels, grades, [0.1]*7, shadow=False)
plt.show()
```

C.
```
grades = [0.02, 0.08, 0.15, 0.25, 0.2, 0.15, 0.15]
labels = ['A+', 'A', 'A-', 'B+', 'B', 'B-', 'C']

plt.pie(labels=labels, x=grades, explode=[0.1]*7)
plt.show()
```

D.
```
grades = [0.02, 0.08, 0.15, 0.25, 0.2, 0.15, 0.15]
labels = ['A+', 'A', 'A-', 'B+', 'B', 'B-', 'C']

plt.pie(grades, [0.1]*7, labels, False)
plt.show()
```

13. What is the output message of the following code?

```
import numpy as np

an_array = np.array([[1, 2],
                     [3, 4]])
result = an_array.sum() - an_array.sum(axis=1)
```

A. 0
B. array([7, 0])
C. array([6, 0])
D. array([7, 3])

14. Consider a regression model $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2 + \hat{\beta}_3 x^3 + \hat{\beta}_4 x^4$, where the model coefficients $\hat{\beta}_i$ are given in a numpy array "b".

```
import numpy as np

b = np.array([20.55, 3.10, 2.71, 0.50, 1.24])
```

Given an array "x" containing five different values for the variable x,

```
x = np.arange(5)
```

which of the following code is correct in calculating the predicted value $\hat{y}$ for each x value? The results should also be given in **an array** "y".

A.
```
y = 0
for n in range(5):
    y = y + b[n]*x**n
```

B.
```
y = 0
for n in range(5):
    y = y + b[n]*x[n]**n
```

C.
```
y = b * x**np.arange(5)
```

D.
```
y = sum(b * x**np.arange(5))
```

15. The table below shows, for credit card holders with one to three cards, the joined probabilities for the number of cards owned (X) and number of credit purchases made in a week (Y).

Number Of Cards (X)	Number of Purchases in Week (Y)				
	0	1	2	3	4
1	0.08	0.13	0.09	0.06	0.03
2	0.03	0.08	0.08	0.09	0.07
3	0.01	0.03	0.06	0.08	0.08

Such a table is given as an array below.

```
import numpy as np

probs = np.array([[0.08, 0.13, 0.09, 0.06, 0.03],
                  [0.03, 0.08, 0.08, 0.09, 0.07],
                  [0.01, 0.03, 0.06, 0.08, 0.08]])
```

Which of the following code is correct in calculating the variance of "the number of purchases in a week"? The variance is represented by the variable **var**.

A.
```
Y = np.arange(5)
```

```
mean = (probs*Y).sum(axis=0)
var = (probs.sum(axis=0)*(Y - mean)**2).sum()
```

B.
```
Y = np.arange(5)

mean = (probs.sum(axis=0)*Y).sum()
var = (probs.sum(axis=0)*(Y - mean)**2).sum()
```

C.
```
Y = np.arange(5)

mean = (probs.sum(axis=0)*Y).sum()
var = (probs.sum()*(Y - mean)**2).sum(axis=0)
```

D.
```
Y = np.arange(5)

mean = (probs.sum(axis=1)*Y).sum()
var = (probs.sum(axis=1)*(Y - mean)**2).sum(axis=0)
```

16. In the case study of "Adult persistence of head-turning asymmetry", it is assumed that the population proportion of adults turning to the right when they are kissing is $p = 0.4$. We are using the binomial distribution "binom" from "scipy.stats" to calculate the probability of having **at least** 10 couples in a 20-case sample turning to the right, which of the following function can be used for the calculation?

A.
```
1 - binom.ppf(9, 20, 0.4)
```

B.
```
binom.cdf(10, 20, 0.4)
```

C.
```
1 - binom.cdf(9, 20, 0.4)
```

D.
```
binom.ppf(10, 20, 0.4)
```

17. Suppose that the amount of a soft drink that goes into a typical 12-ounce can varies from can to can. It follows a normal distribution with the mean value to be 12 ounces and a fixed standard deviation to be 0.05 ounce. How to calculate the probability that the amount of the soft drink is between 11.92 ounces to 12.12 ounces.

A.
```
from scipy.stats import norm

prob = norm.cdf(11.92, 12, 0.05) - (1-norm.cdf(12.12, 12, 0.05))
```

B.
```
from scipy.stats import norm

prob = norm.cdf(12.12, 12, 0.05) - norm.cdf(11.92, 12, 0.05)
```

C.
```
from scipy.stats import norm

prob = 2 - norm.cdf(12.12, 12, 0.05) - norm.cdf(11.92, 12, 0.05))
```

D.
```
from scipy.stats import norm

prob = 1 - norm.cdf(12.12, 12, 0.05) + norm.cdf(11.92, 12, 0.05)
```

18. Given the parameters below,

```
from scipy.stats import norm

alpha = 0.05              # The confidence level 1 - alpha
sigma = 1                 # Population standard deviation
n = 20                    # Sample size n
x_bar = 20                # Sample mean
se = sigma / n**0.5
```

The sample size is “n”, and the sample mean is "x_bar". The population standard deviation is known to be "sigma". How to calculate the lower bound of the confidence interval of the population mean?

A.
```
from scipy.stats import norm

lower = x_bar - norm.ppf(1-alpha/2)*se
```

B.
```
from scipy.stats import norm

lower = x_bar - norm.ppf((1-alpha)/2)*se
```

C.
```
from scipy.stats import norm

lower = x_bar + norm.ppf(-alpha/2)*se
```

D.
```
from scipy.stats import norm

lower = x_bar + norm.ppf((1-alpha)/2)*se
```

19. Given a data table "wage_data" below,

	wage	educ	exper	female	married
0	3.10	11.0	2.0	1.0	0.0
1	3.24	12.0	22.0	1.0	1.0
2	3.00	11.0	2.0	0.0	0.0
3	6.00	8.0	44.0	0.0	1.0
4	5.30	12.0	7.0	0.0	1.0

Which of the following code gives the median of the attribute “wage”?

A.
```
wage_summary = wage_data.describe()
print(wage_summary['wage']['median'])
```

B.
```
wage_summary = wage_data.describe()
print(wage_summary['median']['wage'])
```

C.
```
wage_summary = wage_data.describe()
print((wage_summary['wage']['min']+wage_summary['wage']['max'])/2)
```

D.
```
wage_summary = wage_data.describe()
print((wage_summary['wage']['50%']))
```

20. Which of the following models can be viewed as a linear regression model?

(1) $\hat{y} = \beta_0 + \beta_1 x$

(2) $\sqrt{\hat{y}} = \frac{\beta_1}{x} + \beta_2 x^2$

(3) $\hat{y} = \beta_0 + \beta_1^x$

(4) $\log \hat{y} = \beta_0 + \beta_1 \log x$

A. (1), (2), and (3)
B. (1), (2), and (4)
C. (1) and (2)
D. Only (1)

21. Which of the following statements on multiple regression model is correct?

A. If the slope parameter β_j is positive, then x_j is positively correlated with the dependent variable y.
B. If x_j is positively correlated with the dependent variable y, then the slope parameter β_j is positive.
C. If β_j is larger than β_k, then we can conclude that x_j has a greater impact on the dependent variable y, compared with x_k.
D. All above are incorrect

PART II. QUANTITATIVE QUESTIONS (37 marks)

We toss a dice 65 times, and the results are given in the following table.

Tossed Number	Number of Appearances
1	12
2	7
3	11
4	9
5	15
6	11

(a) What is the sample mean of the tossed number?

$E[X] = \frac{12}{65} \times 1 + \frac{7}{65} \times 2 + \frac{11}{65} \times 3 + \frac{9}{65} \times 4 + \frac{15}{65} \times 5 + \frac{11}{65} \times 6 = 3.6307$	5 marks

(b) What is the sample proportion of the event that the tossed number is 4?

$\hat{p} = \frac{9}{65} = 0.138$	5 marks

(c) Write down the Python code to calculate confidence interval of the population proportion of "4" as the tossed number.

`# Use ps to denote the sample proportion of 4`
`# Use lower to represent the lower bound`
`# Use upper to represent the upper bound`
 	
`from scipy.stats import norm`
 	
`n = 65`
`alpha = 0.05`
 	
`lower = ps-norm.ppf(1-alpha/2)*(ps*(1-ps)/n)**0.5`
`upper = ps+norm.ppf(1-alpha/2)*(ps*(1-ps)/n)**0.5`
	8 marks

(d) Based on the given data, can we conclude that the proportion of having "4" as the tossed number is different from 1/6? Answer this question using the hypothesis testing (P-value), with the significance level to be $\alpha = 0.05$

The null and alternative hypotheses:

$H_0: p = 1/6$ $H_a: p \neq 1/6$	5 marks

Python code to calculate the P-value:

`# Use ps to denote the sample proportion of 4`
`# Use p_value to represent the P-value`
 	
`from scipy.stats import norm`
 	
`n = 65`
`z_value = (ps - (1/6)) / ((1/6)*(1-1/6)/n)**0.5`
 	
`p_value = 2*(1 - norm.cdf(abs(z_value)))`
	9 marks

What is the interpretation of the P-value (how the P-value affects your conclusion)?

If P-value is larger than alpha, there is insufficient evidence to reject the null hypothesis under the given significance level. If P-value is lower than alpha, we reject the null hypothesis in favor of the alternative hypothesis, which implies that the proportion of having "4" as the tossed number is different from 1/6.	5 marks

Total Score for Part II	