IT3030 midterm exam#1 3/27/2018 ID:_____ Name:_____

An OPEN-book test.

Part 1 – Choose (circle) the BEST answer. (40%)

1. Age is classified as

A. Nominal data **B. Ordinal data** C. Categorical data

2. Which of the followings can you use to describe categorical variable?

A. Frequency (%) B. Mean (SD) C. Median (IQR)

3. From the list, select one that DOES NOT describe central tendency.

A. Mean B. Median C. Inter-quartile range (IQR) **D. Standard deviation (SD)**

4. In standard normal distribution, mean, median and mode are NOT equal.

A. True **B. False**

5. Which of the followings best describes the MODE?

A. The value that has the highest frequency.

B. The average calculated by adding up all the values and dividing by the number of entries.

C. The largest value.

D. The value that half of the entries are below and half of the entries are above.

6. Which of the followings defines the MEDIAN?

A. The value that half of the entries are below and half of the entries are above.

B. The value that has the highest frequency.

C. The largest value.

D. The average calculated by adding up all the values and dividing by the number of entries.

7. What is the mode and the mean for the following set of numbers? {4, 9, 8, 2, 16, 4, 4, 8, 9, 6}

A. Mean = 7, mode = 8 **B. Mean = 7, mode = 4**

C. Mean = 6, mode = 8 D. Mean = 8, mode = 9

8. In descriptive statistics our main objective is to

A. Describe the population **B. Describe the data we collected**

C. Compute an average D. Infer something about the population

2. (20%) Given the frequency table below describing the scores for a group of students. (a) Graph the histogram (using absolute frequencies) of these scores. (b) Determine the mean value of the scores. (c) Is your mean value appropriate for representing the central tendency of these scores? Yes or no? State your reason(s). (d) Determine the standard deviation of these scores.

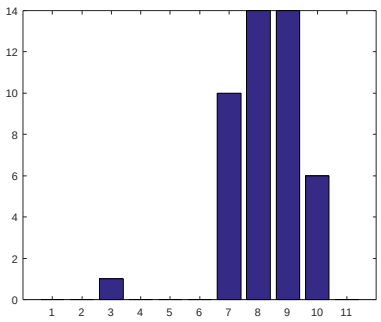
Score	Number of students
0-9	0
10-19	0
20-29	1
30-39	0
40-49	0
50-59	0
60-69	10
70-79	14
80-89	14
90-99	6
100	0

```
>> f=[0 0 1 0 0 0 10 14 14 6 0];
>>bar(f)

>> m=[4.5 14.5 24.5 34.5 44.5 54.5 64.5 74.5 84.5 94.5 100];
>> mu=sum(f.*m)/sum(f)
mu =      76.9444
```

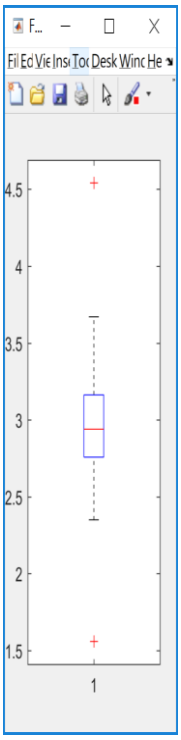
Looks like the 'mode' is near the '70-79' group, which fits the mean value we obtained here.

```
>> sqrt((sum(((m-mu).^2).*f)/(sum(f)-1)))
ans =
    12.6411
```



3. (20%) Given the following FEV1 data for 13 subjects. (a) Determine the Q1, Q2 and Q3, respectively. (b) Graph the box plot (no whiskers yet). (3) Taking 2 times of the box height for marking the 'whiskers', determine the two boundaries and whether you have any outliers for these 13 observations. (d) Adding the whiskers as well as all these 13 values to your box plot. Are your Q1, Q2, Q3 coinciding with these 13 values? Yes or No? Why?

Subject	FEV1
1	3.11
2	**1.56**
3	2.94
4	3.01
5	2.88
6	**4.54**
7	2.78
8	2.69
9	3.67
10	3.32
11	2.87
12	3.09
13	2.35




```
>> boxplot(x,'whisker',2)  【default whisker is 1.5】

>> quantile(x,0.25)
ans =      2.7575
>> quantile(x,0.5)
ans =      2.9400
>> median(x)
ans =      2.9400
>> quantile(x,0.75)
ans =      3.1625
>>
>> lower=quantile(x,0.25)-2*(quantile(x,0.75)-quantile(x,0.25))
lower =      1.9475    Outlier = 1.56 (2nd entry)
>> higher=quantile(x,0.75)+2*(quantile(x,0.75)-quantile(x,0.25))
higher =      3.9725    Outlier = 4.54 (6th entry)
```

4. (20%) Following the same example of infant deaths in US we had in class, subject to the following new statistics:

(1) Live Births, Infant Deaths for Colorado are 54,345 and 509, respectively.

(2) Live Births, Infant Deaths for Louisiana are 88,113 and 1,024, respectively.

(3) Assuming the compositions of black, white and others of live birth and infant death remain the same as shown in class.

(4) Assuming the US population compositions and numbers are the same.

(a) Complete the empty cells in the table below. (b) Determine the adjusted death rate for both states by direct method. (c) Determine the adjusted death rate for both states by indirect method.

	Colorado			Louisiana		
Race	Live birth	Infant deaths	Rate per 1000	Live birth	Infant deaths	Rate per 1000
Black	3,198	50	15.7	35,344	617	17.4
White	49,292	453	9.2	50,925	404	7.9
Other	1,855	6	3.1	1,844	4	1.9
total	54,345	509	9.4	88,113	1024	11.6

(Take integer for counts. Take one digit after decimal point for rates.)

Race	Live Births	Rate per 1000	Expected Deaths	Rate per 1000	Expected Deaths
Black	641,567	15.7	10,072.60	17.4	11,163.27
White	2,992,488	9.2	27,530.89	7.9	23,640.66
Other	175,339	3.1	543.55	1.9	333.14
Total	3,809,394	9.4	38,147.04	11.6	35,137.07
	adjusted	10.01	adjusted	9.22	

Adjusted Colorado = (38,147.04/3,809,394)*1000 = 10.01

Adjusted Louisiana = **(**35,137.07/3,809,394)*1000 = 9.22

Race	Rate per 1000	Live Births	Expected Deaths	Live Births	Expected Deaths
Black	17.9	3,198	57.24	35,344	632.66
White	8.6	49,292	423.91	50,925	437.96
Other	6.5	1,855	12.06	1,844	11.99
Total	10.1	54,345	493.21	88,113	1,082.60
		adjusted	10.42	adjusted	9.55

Adjusted Colorado = 10.1*(509/493.21) = 10.42

Adjusted Louisiana = 10.1*(1024/1,082.60) = 9.55