Quiz 1

- There are 6 problems total worth 35 points as shown in each question.
- You must not communicate with other students during this test.
- No books, notes allowed.
- No other electronic device except calculators are allowed. You cannot use your mobile as calculators.
- This is a 45 minute exam.
- Do not turn this page until instructed to.
- There are several different versions of this exam.

1. Fill in your information:

Full Name: ______________________________

NetID: ______________________________

1/1. (5 points) The least and greatest number in a list of 7 integer numbers are 2 and 20 respectively. The median and mode of list are 6 and 3 respectively. Find out which of the following options can be mean of the list.

A. 4

B. 7

C. 6.85

D. 6.71

Note there may be more than one correct option. You have to identify all to get full credit.

Solution. $2, 3, 3, 6, x, y, 20$

x and y has to be at least 7 and 8 respectively.

B is the correct answer

1/2. (5 points) The least and greatest number in a list of 7 integer numbers are 1 and 20 respectively. The median and mode of list are 9 and 2 respectively. Find out which of the following options can be mean of the list.

A. 6

B. 7.71

C. 7.86

D. 8

Note there may be more than one correct option. You have to identify all to get full credit.

Solution. $1, 2, 2, 9, x, y, 20$

x and y has to be at least 10 and 11 respectively.

C and D both are the correct answer

1/3. (5 points) The least number in a list of 5 integer numbers is 2. The median and mode of list are 5 and 7 respectively. Find out which of the following options can be mean of the list.

A. 4.6

B. 5

C. 4.8

D. 5.2

Note there may be more than one correct option. You have to identify all to get full credit.

Solution. $2, x, 5, 7, 7$

x can be either 3 or 4

B and C are the correct answer

1/4. (5 points) The least and greatest number in a list of 7 integer numbers are 2 and 20 , respectively. The median and mode of list are 6 and 7 respectively. Find out which of the following options can be mean of the list.

A. 4

B. 7

C. 6.85

D. 7.14

Note there may be more than one correct option. You have to identify all to get full credit.

Solution. $2, x, y, 6, 7, 7, 20$

x and y must be one of 3 pairs - (3,4) , (3,5), (4,5)

B,D are the correct answer

2/1. (5 points)

Consider the following data on age distribution of a population:

age	frequency
1-5	200
6-15	450
16-20	300
21-50	1500
51-80	700
81-110	44

A. Compute the approximate median age for the given population.

B. Identify the case that results in maximum error in the calculation of the median; assume that the frequencies for each interval is the same as given in the table.

Solution.

A. 33.5

B. When all 1500 people in the median interval have age 50. 50 - 33.5 = 16.5

2/2. (5 points)

Consider the following data on age distribution of a population:

age	frequency
1-5	200
6-15	450
16-20	300
21-50	1500
51-80	700
81-110	444

A. Compute the approximate median age for the given population.

B. Identify the case that results in maximum error in the calculation of the median; assume that the frequencies for each interval is the same as given in the table.

Solution.

A. 37.38

B. When all 1500 people in the median interval have age 21. 37.38 - 21 = 16.38

3/1. (5 points) Consider the following data tuples : $(0, 1), (6, 6), (8, 7)$
Compute the following tuple (L_1, L_2, L_∞) for each pair of data tuples.
Solution.

A. $(0, 1), (6, 6) \rightarrow (11, 7.81, 6)$

B. $(0, 1), (8, 7) \rightarrow (14, 10, 8)$

C. $(6, 6), (8, 7) \rightarrow (3, 2.24, 2)$

3/2. (5 points) Consider the following data tuples : $(1, 2), (7, 7), (9, 8)$
Compute the following tuple (L_1, L_2, L_∞) for each pair of data tuples.
Solution.

A. $(1, 2), (7, 7) \rightarrow (11, 7.81, 6)$

B. $(1, 2), (9, 8) \rightarrow (14, 10, 8)$

C. $(7, 7), (9, 8) \rightarrow (3, 2.24, 2)$

3/3. (5 points) Consider the following data tuples : $(2, 3), (8, 8), (10, 9)$
Compute the following tuple (L_1, L_2, L_∞) for each pair of data tuples.
Solution.

A. $(2, 3), (8, 8) \rightarrow (11, 7.81, 6)$

B. $(2, 3), (10, 9) \rightarrow (14, 10, 8)$

C. $(8, 8), (10, 9) \rightarrow (3, 2.24, 2)$

3/4. (5 points) Consider the following data tuples : $(3, 4), (9, 9), (11, 10)$
Compute the following tuple (L_1, L_2, L_∞) for each pair of data tuples.
Solution.

A. $(3, 4), (9, 9) \rightarrow (11, 7.81, 6)$

B. $(3, 4), (9, 9) \rightarrow (14, 10, 8)$

C. $(9, 9), (11, 10) \rightarrow (3, 2.24, 2)$

4/1. (5 points) Consider following data on 2 persons Jack and Jill.
Height and Weight are given in ft and lbs respectively.

Name	Age	Height	Weight
Jack	24	7	210
Jill	14	4	140

Age , Height and Weight are **ordinal attributes** with following interval states.

A. **Age** : [3-18], [19-40], [41-80]

B. **Height** : [3-5], [6-9]

C. **Weight** : [70-110], [110-150], [150-190], [190-230]

Compute the Manhattan distance between 2 persons. Show the steps of calculation.
Solution. Jack : (2 , 2, 4)
Jill : (1, 1, 2)
Each ordinal attributes has different number of states. So they need to be normalized.
Jack : $(\frac{1}{2}, 1, 1)$
Jill : $(0, 0, \frac{1}{3})$
Manhattan Distance : $\frac{13}{6}$

4/2. (5 points) Consider following data on 2 persons Jack and Jill.
Height and Weight are given in ft and lbs respectively.

Name	Age	Height	Weight
Jack	27	7.5	220
Jill	12	4.2	120

Age , Height and Weight are **ordinal attributes** with following interval states.

A. **Age** : [3-18], [19-40], [41-80]

B. **Height** : [3-5], [6-9]

C. **Weight** : [70-110], [110-150], [150-190], [190-230]

Compute the Manhattan distance between 2 persons. Show the steps of calculation.
Solution. Jack : (2 , 2, 4)
Jill : (1, 1, 2)
Each ordinal attributes has different number of states.So they need to be normalized.
Jack : $(\frac{1}{2}, 1, 1)$
Jill : $(0, 0, \frac{1}{3})$
Manhattan Distance : $\frac{13}{6}$

4/3. (5 points) Consider following data on 2 persons Jack and Jill.
Height and Weight are given in ft and lbs respectively.

Name	Age	Height	Weight
Jack	38	8	225
Jill	4	3	112

Age , Height and Weight are **ordinal attributes** with following interval states.

A. **Age** : [3-18], [19-40], [41-80]

B. **Height** : [3-5], [6-9]

C. **Weight** : [70-110], [110-150], [150-190], [190-230]

Compute the Manhattan distance between 2 persons. Show the steps of calculation.
Solution. Jack : (2 , 2, 4)
Jill : (1, 1, 2)
Each ordinal attributes has different number of states.So they need to be normalized.
Jack : $(\frac{1}{2}, 1, 1)$
Jill : $(0, 0, \frac{1}{3})$
Manhattan Distance : $\frac{13}{6}$

5/1. (5 points) Consider 2 students A and B. They have both taken CS 241 course but in different semesters. The class size was same in both the semesters i.e 100. Now A's total weighted score at the end of semester was 80. On the other hand, B scored 60.

Who did better relative to the class? Explain.
Here are some statistics on the 2 class :

Class	Mean Score	Standard Deviation
A's Class	70	10
B's Class	40	10

Solution. $z_A = \frac{80-70}{10} = 1.0$
$z_B = \frac{60-40}{10} = 2.0$
So B did better.

5/2. (5 points) Consider 2 students A and B. They have both taken CS 241 course but in different semesters. The class size was same in both the semesters i.e 100. Now A's total weighted score at the end of semester was 60. On the other hand, B scored 90.

Who did better relative to the class? Explain.
Here are some statistics on the 2 class :

Class	Mean Score	Standard Deviation
A's Class	50	10
B's Class	90	5

Solution. $z_A = \frac{60-50}{10} = 1.0$
$z_B = \frac{90-90}{5} = 0$
So A did better.

5/3. (5 points) Consider 2 students A and B. They have both taken CS 241 course but in different semesters. The class size was same in both the semesters i.e 100. Now A's total weighted score at the end of semester was 70. B also scored 70.

Who did better relative to the class? Explain.
Here are some statistics on the 2 class :

Class	Mean Score	Standard Deviation
A's Class	50	20
B's Class	50	10

Solution. $z_A = \frac{70-50}{20} = 1.0$
$z_B = \frac{70-50}{10} = 2.0$
So B did better.

6/1. (10 points) Consider a dataset $x_1, x_2, ..., x_n$

Mean Absolute Deviation(**MAD**) of set is given as $\frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|$

Standard Deviation(σ) of set is given as $\sqrt{\frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|^2}$

where $\bar{x}$ is the mean of dataset.

Find out an inequality relationship between MAD and σ

Hint : Jensen's inequality states that for variable x,

$$f(E(x)) \leq E(f(x)) \quad \forall f(x) \text{is a convex function}$$

where $E(x)$ signifies the expectation/mean of the variable x

A real-valued function defined on an interval is called convex if the line segment between any two points on the graph of the function lies above or on the graph.Some examples of convex functions are x^2, x^4, x^8

Solution. Let variable $X = |x - \bar{x}|$

So $X_i = |x_i - \bar{x}|$

$$E(X) = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}| = MAD$$

Let $f(y) = y^2$. So ,

$$f(E(X)) = (MAD)^2$$

$$E(f(X)) = E(|x - \bar{x}|^2) = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|^2 = \sigma^2$$

Using Jensen's inequality, $MAD^2 \leq \sigma^2$ Relation : MAD $\leq \sigma$
