The LNM Institute of Information Technology, Jaipur

CSE327 Introduction to Data Science

Quiz-I **Thursday, Sep 19, 2018**

Time: 45 Minutes Maximum Marks: 40

Name: _____ Roll No:_____

Read the Instructions Carefully:
All are objective questions with 4 choices. There may be more than one correct answers. Each question carry two marks. You need to select all the correct answers to get full marks. For example, if a question has 2 correct choices and you did only one correctly you will get 1 mark. If out of your choices even one choice is wrong then there won't be any marks given (even if there are correct choices that you have made). So be careful when you are finding and marking the correct choices. **Encircle all the correct answer/s. Any other marking will be considered as invalid**. Do not write anything in the answer sheet except the answer. No justification is required. **There is no negative marking for wrong answer/s.**

1. Which of the following does not make sense at all:
   (a) Mode of a list of nominal values
   (b) Mean of a list of ordinal values
   (c) Mean of a list of nominal values
   (d) Median of a list of ordinal values

2. Pick the correct statement:
   (a) Histogram is for quantitative data
   (b) Bar Chart is for categorical data
   (c) Histogram is for categorical data
   (d) Bar Chart is for quantitative data

3. At an Institute 70 students are raising money for Kerala flood relief fund. The average amount raised is ₹ 100. What is the largest possible number of children who could have raised ₹ 1000 or more?
   (a) 6
   (b) 7
   (c) 8
   (d) 9

4. A list of numbers has been converted to standard units. See the table given below. Two of the cells in the table are unknown. Find it!

value	standard unit
54	2.8
A	-1.2
42	0.4
49	B

   (a) A=34 and B=1.8
   (b) A=40 and B=1.8
   (c) A=34 and B=1.5
   (d) A=40 and B=1.5

5. A simple random sample of 1900 households (How many people are there in a house) is taken in Jaipur. The average household size in the sample is 2.2 people, with an SD of 2 people. Pick the correct (more appropriate) statement(s):
   (a) The average household size in the sample is known to be 2.2
   (b) The average household size in the sample is estimated to be 2.2
   (c) The average household size in the city is known to be 2.2
   (d) The average household size in the city is estimated to be 2.2

6. Given a mark list of students of IDS course that has an average of 50 and an SD of 10. What proportion of students would have got marks less than 90?

   (a) 3/4
   (b) 1/4
   (c) 1/16
   (d) 15/16

7. Here is a stem and leaf plot of the heights (in centimeters) of a group of men.
   16|56778
   17|0012234444
   17|55666788889
   18|0111234
   18|56
   How many men are 178cms tall and what is the range of the heights (i.e., difference between the smallest and the tallest height)?

   (a) 3 and 20cms
   (b) 3 and 21cms
   (c) 4 and 21cms
   (d) 4 and 20cms

8. A survey company took a simple random sample of 275 rental flats out of all the rental flats in Bengaluru. The average monthly rent of the sampled units was ₹15000 and the SD was ₹9000. There were 964 people living in the sampled flats, and there were 120 children among the these 964 people. Pick the correct statements:

   (a) Average monthly rent of flats in Bengaluru is estimated to be ₹15000
   (b) This distribution will not be a normal distribution
   (c) 964 people is a good simple random sample to calculate the percentage of children living in the flats in Bengaluru
   (d) 964 people is not a good simple random sample to calculate the percentage of children living in the flats in Bengaluru

9. A simple random sample of 1900 households (How many people are there in a house) is taken in Jaipur. The average household size in the sample is 2.2 people, with an SD of 2 people. The distribution of household size is

   (a) Perfectly Normal
   (b) Approximately Normal
   (c) Is not Normal at all
   (d) We can't say. Need more data

10. Given is a list of numbers that has an average of 50 and an SD of 10. Is there a way to transform the list somehow so that the list has the SD of 30 but leaving the average the same?

    (a) Yes
    (b) No
    (c) Not possible with this information
    (d) Possible only if all are positive integers

11. Marks in a test are between 0 and 100. The average is 85 and the SD is 12. Based on this data, can we use the normal approximation to estimate percentiles of the distribution? If not, what shall we use to get a good estimation?

    (a) Normal approximation is enough
    (b) Normal approximation cannot be used we should use Markov's inequality
    (c) Normal approximation cannot be used we should use Chebechev's inequality
    (d) Not possible to decide we need more data

12. Piechart is mostly meant for

    (a) Ordinal variable
    (b) Nominal variable
    (c) Continuous variable
    (d) Discrete variable

13. A simple random sample of 1900 households (How many people are there in a house) is taken in Jaipur. The average household size in the sample is 2.2 people, with an SD of 2 people. Now how about this statement: *Approximately 95% of the households had sizes in the range 2.07 to 2.33 people*

    (a) True statement

    (b) False statement

    (c) Valid statement but need to check the truthfulness with more data

    (d) Invalid statement

14. An Institute records the distance traveled from home to work of all of its 80 employees. By mistake, the largest commute distance is recorded as 65kms instead of 6.5kms. By how many miles does this mistake change the median of the commute distances? And how does the mean change?

    (a) Median by 16.5kms and Mean by 0.73kms

    (b) Median by 16.5kms and Mean by 0.72kms

    (c) Median by 0kms and Mean by 0.73kms

    (d) Median by 0kms and Mean by 0.72kms

15. For the income data distribution given below, find the *40th percentile* in thousands of dollars. In the calculation if you get decimals please round it off to two decimal places.

0 – $10000	20%
$10001 – $25000	28%
$25001 – $50000	27%
$50001 – $100000	18%
$100000 – $150000	7%

    (a) $20710

    (b) $20720

    (c) $20170

    (d) $20180

16. For the income data distribution given below, find the *inter-quartile range*, that is, the distance between the 75th percentile and the 25th percentile. In the calculation if you get decimals please round it off to two decimal places.

0 – $10000	20%
$10001 – $25000	28%
$25001 – $50000	27%
$50001 – $100000	18%
$100000 – $150000	7%

    (a) $37230

    (b) $37320

    (c) $37310

    (d) $37240

For the following four MCQs use the below information:

We are going to test whether or not a random number generator is producing the digit 0 *in the correct proportion, i.e.,* 1/10. *For this we run the generator 5,000 times. We assume that the runs are mutually independent and that each has the same probability p of producing* 0.

We would like to construct a test that has a significance level of approximately 1%. Towards this answer the following MCQs.

17. The distribution of 0 follows what kind of a distribution (precisely)?

    (a) Normal

    (b) Poisson

    (c) Bernoulli

    (d) Binomial

18. What is the mean of the above distribution?

    (a) 500

    (b) 50

    (c) 5000

    (d) None of the above

19. What is the SD of the above distribution?

    (a) 21.12

    (b) 21.21

    (c) 22.36

    (d) 22.63

20. In order to test the hypothesis which of the following will be the most appropriate:

(a) $H_0 : p = 0.1$ and $H_A : p \neq 0.1$

(b) $H_0 : p = 0.1$ and $H_A : p \leq 0.1$

(c) $H_0 : p = 0.1$ and $H_A : p \geq 0.1$

(d) $H_0 : p \leq 0.1$ and $H_A : p \geq 0.1$

Answer Key for Exam A

1. Which of the following does not make sense at all:

   (a) Mode of a list of nominal values

   (b) Mean of a list of ordinal values

   [(c)] Mean of a list of nominal values

   [(d)] Median of a list of ordinal values

2. Pick the correct statement:

   [(a)] Histogram is for quantitative data

   [(b)] Bar Chart is for categorical data

   (c) Histogram is for categorical data

   (d) Bar Chart is for quantitative data

3. At an Institute 70 students are raising money for Kerala flood relief fund. The average amount raised is ₹ 100. What is the largest possible number of children who could have raised ₹ 1000 or more?

   (a) 6

   [(b)] 7

   (c) 8

   (d) 9

4. A list of numbers has been converted to standard units. See the table given below. Two of the cells in the table are unknown. Find it!

value	standard unit
54	2.8
A	-1.2
42	0.4
49	B

   [(a)] A=34 and B=1.8

   (b) A=40 and B=1.8

   (c) A=34 and B=1.5

   (d) A=40 and B=1.5

5. A simple random sample of 1900 households (How many people are there in a house) is taken in Jaipur. The average household size in the sample is 2.2 people, with an SD of 2 people. Pick the correct (more appropriate) statement(s):

   [(a)] The average household size in the sample is known to be 2.2

   (b) The average household size in the sample is estimated to be 2.2

   (c) The average household size in the city is known to be 2.2

   [(d)] The average household size in the city is estimated to be 2.2

6. Given a mark list of students of IDS course that has an average of 50 and an SD of 10. What proportion of students would have got marks less than 90?

   (a) 3/4
   (b) 1/4
   (c) 1/16
   [(d)] 15/16

7. Here is a stem and leaf plot of the heights (in centimeters) of a group of men.
   16|56778
   17|0012234444
   17|55666788889
   18|0111234
   18|56
   How many men are 178cms tall and what is the range of the heights (i.e., difference between the smallest and the tallest height)?

   (a) 3 and 20cms
   (b) 3 and 21cms
   [(c)] 4 and 21cms
   (d) 4 and 20cms

8. A survey company took a simple random sample of 275 rental flats out of all the rental flats in Bengaluru. The average monthly rent of the sampled units was ₹15000 and the SD was ₹9000. There were 964 people living in the sampled flats, and there were 120 children among the these 964 people. Pick the correct statements:

   [(a)] Average monthly rent of flats in Bengaluru is estimated to be ₹15000
   [(b)] This distribution will not be a normal distribution
   (c) 964 people is a good simple random sample to calculate the percentage of children living in the flats in Bengaluru
   [(d)] 964 people is not a good simple random sample to calculate the percentage of children living in the flats in Bengaluru

9. A simple random sample of 1900 households (How many people are there in a house) is taken in Jaipur. The average household size in the sample is 2.2 people, with an SD of 2 people. The distribution of household size is

   (a) Perfectly Normal
   (b) Approximately Normal
   [(c)] Is not Normal at all
   (d) We can't say. Need more data

10. Given is a list of numbers that has an average of 50 and an SD of 10. Is there a way to transform the list somehow so that the list has the SD of 30 but leaving the average the same?

    [(a)] Yes
    (b) No
    (c) Not possible with this information
    (d) Possible only if all are positive integers

11. Marks in a test are between 0 and 100. The average is 85 and the SD is 12. Based on this data, can we use the normal approximation to estimate percentiles of the distribution? If not, what shall we use to get a good estimation?

    (a) Normal approximation is enough

    (b) Normal approximation cannot be used we should use Markov's inequality

    [(c)] Normal approximation cannot be used we should use Chebechev's inequality

    (d) Not possible to decide we need more data

12. Piechart is mostly meant for

    (a) Ordinal variable

    [(b)] Nominal variable

    (c) Continuous variable

    (d) Discrete variable

13. A simple random sample of 1900 households (How many people are there in a house) is taken in Jaipur. The average household size in the sample is 2.2 people, with an SD of 2 people. Now how about this statement: *Approximately 95% of the households had sizes in the range 2.07 to 2.33 people*

    (a) True statement

    (b) False statement

    (c) Valid statement but need to check the truthfulness with more data

    [(d)] Invalid statement

14. An Institute records the distance traveled from home to work of all of its 80 employees. By mistake, the largest commute distance is recorded as 65kms instead of 6.5kms. By how many miles does this mistake change the median of the commute distances? And how does the mean change?

    (a) Median by 16.5kms and Mean by 0.73kms

    (b) Median by 16.5kms and Mean by 0.72kms

    [(c)] Median by 0kms and Mean by 0.73kms

    (d) Median by 0kms and Mean by 0.72kms

15. For the income data distribution given below, find the *40th percentile* in thousands of dollars. In the calculation if you get decimals please round it off to two decimal places.

0 – \$10000	20%
\$10001 – \$25000	28%
\$25001 – \$50000	27%
\$50001 – \$100000	18%
\$100000 – \$150000	7%

    [(a)] \$20710

    (b) \$20720

    (c) \$20170

    (d) \$20180

16. For the income data distribution given below, find the *inter-quartile range*, that is, the distance between the 75th percentile and the 25th percentile. In the calculation if you get decimals please round it off to two decimal places.

0 – \$10000	20%
\$10001 – \$25000	28%
\$25001 – \$50000	27%
\$50001 – \$100000	18%
\$100000 – \$150000	7%

(a) \$37230
(b) \$37320
(c) \$37310
(d) \$37240

For the following four MCQs use the below information:

We are going to test whether or not a random number generator is producing the digit 0 *in the correct proportion, i.e.,* 1/10. *For this we run the generator 5,000 times. We assume that the runs are mutually independent and that each has the same probability* p *of producing* 0.

We would like to construct a test that has a significance level of approximately 1%. Towards this answer the following MCQs.

17. The distribution of 0 follows what kind of a distribution (precisely)?

(a) Normal
(b) Poisson
(c) Bernoulli
(d) Binomial

18. What is the mean of the above distribution?

(a) 500
(b) 50
(c) 5000
(d) None of the above

19. What is the SD of the above distribution?

(a) 21.12
(b) 21.21
(c) 22.36
(d) 22.63

20. In order to test the hypothesis which of the following will be the most appropriate:

(a) $H_0 : p = 0.1$ and $H_A : p \neq 0.1$
(b) $H_0 : p = 0.1$ and $H_A : p \leq 0.1$
(c) $H_0 : p = 0.1$ and $H_A : p \geq 0.1$
(d) $H_0 : p \leq 0.1$ and $H_A : p \geq 0.1$

The LNM Institute of Information Technology, Jaipur

CSE327 Introduction to Data Science

Quiz-I **Thursday, Sep 19, 2018**

Time: 45 Minutes Maximum Marks: 40

Name: _____ Roll No:_____

Read the Instructions Carefully:
All are objective questions with 4 choices. There may be more than one correct answers. Each question carry two marks. You need to select all the correct answers to get full marks. For example, if a question has 2 correct choices and you did only one correctly you will get 1 mark. If out of your choices even one choice is wrong then there won't be any marks given (even if there are correct choices that you have made). So be careful when you are finding and marking the correct choices. **Encircle all the correct answer/s. Any other marking will be considered as invalid**. Do not write anything in the answer sheet except the answer. No justification is required. **There is no negative marking for wrong answer/s.**

1. For the income data distribution given below, find the *inter-quartile range*, that is, the distance between the 75th percentile and the 25th percentile. In the calculation if you get decimals please round it off to two decimal places.

0 – $10000	20%
$10001 – $25000	28%
$25001 – $50000	27%
$50001 – $100000	18%
$100000 – $150000	7%

   (a) $37230
   (b) $37320
   (c) $37310
   (d) $37240

2. Marks in a test are between 0 and 100. The average is 85 and the SD is 12. Based on this data, can we use the normal approximation to estimate percentiles of the distribution? If not, what shall we use to get a good estimation?
   (a) Normal approximation is enough
   (b) Normal approximation cannot be used we should use Markov's inequality
   (c) Normal approximation cannot be used we should use Chebechev's inequality
   (d) Not possible to decide we need more data

3. A simple random sample of 1900 households (How many people are there in a house) is taken in Jaipur. The average household size in the sample is 2.2 people, with an SD of 2 people. Now how about this statement: *Approximately 95% of the households had sizes in the range 2.07 to 2.33 people*
   (a) True statement
   (b) False statement
   (c) Valid statement but need to check the truthfulness with more data
   (d) Invalid statement

4. An Institute records the distance traveled from home to work of all of its 80 employees. By mistake, the largest commute distance is recorded as 65kms instead of 6.5kms. By how many miles does this mistake change the median of the commute distances? And how does the mean change?
   (a) Median by 16.5kms and Mean by 0.73kms
   (b) Median by 16.5kms and Mean by 0.72kms
   (c) Median by 0kms and Mean by 0.73kms
   (d) Median by 0kms and Mean by 0.72kms

For the following four MCQs use the below information:

We are going to test whether or not a random number generator is producing the digit 0 *in the correct proportion, i.e.,* 1/10. *For this we run the generator 5,000 times. We assume that the runs are mutually independent and that each has the same probability* p *of producing* 0.

We would like to construct a test that has a significance level of approximately 1%. Towards this answer the following MCQs.

5. The distribution of 0 follows what kind of a distribution (precisely)?
   (a) Normal
   (b) Poisson
   (c) Bernoulli
   (d) Binomial

6. What is the mean of the above distribution?
   (a) 500
   (b) 50
   (c) 5000
   (d) None of the above

7. What is the SD of the above distribution?
   (a) 21.12
   (b) 21.21
   (c) 22.36
   (d) 22.63

8. In order to test the hypothesis which of the following will be the most appropriate:
   (a) $H_0 : p = 0.1$ and $H_A : p \neq 0.1$
   (b) $H_0 : p = 0.1$ and $H_A : p \leq 0.1$
   (c) $H_0 : p = 0.1$ and $H_A : p \geq 0.1$
   (d) $H_0 : p \leq 0.1$ and $H_A : p \geq 0.1$

9. Here is a stem and leaf plot of the heights (in centimeters) of a group of men.
   16|56778
   17|0012234444
   17|55666788889
   18|0111234
   18|56
   How many men are 178cms tall and what is the range of the heights (i.e., difference between the smallest and the tallest height)?
   (a) 3 and 20cms
   (b) 3 and 21cms
   (c) 4 and 21cms
   (d) 4 and 20cms

10. For the income data distribution given below, find the *40th percentile* in thousands of dollars. In the calculation if you get decimals please round it off to two decimal places.

0 – \$10000	20%
\$10001 – \$25000	28%
\$25001 – \$50000	27%
\$50001 – \$100000	18%
\$100000 – \$150000	7%

   (a) \$20710
   (b) \$20720
   (c) \$20170
   (d) \$20180

11. Pick the correct statement:
   (a) Histogram is for quantitative data
   (b) Bar Chart is for categorical data
   (c) Histogram is for categorical data
   (d) Bar Chart is for quantitative data

12. Given is a list of numbers that has an average of 50 and an SD of 10. Is there a way to transform the list somehow so that the list has the SD of 30 but leaving the average the same?
   (a) Yes
   (b) No
   (c) Not possible with this information
   (d) Possible only if all are positive integers

13. A simple random sample of 1900 households (How many people are there in a house) is taken in Jaipur. The average household size in the sample is 2.2 people, with an SD of 2 people. The distribution of household size is

    (a) Perfectly Normal

    (b) Approximately Normal

    (c) Is not Normal at all

    (d) We can't say. Need more data

14. Which of the following does not make sense at all:

    (a) Mode of a list of nominal values

    (b) Mean of a list of ordinal values

    (c) Mean of a list of nominal values

    (d) Median of a list of ordinal values

15. Given a mark list of students of IDS course that has an average of 50 and an SD of 10. What proportion of students would have got marks less than 90?

    (a) 3/4

    (b) 1/4

    (c) 1/16

    (d) 15/16

16. A simple random sample of 1900 households (How many people are there in a house) is taken in Jaipur. The average household size in the sample is 2.2 people, with an SD of 2 people. Pick the correct (more appropriate) statement(s):

    (a) The average household size in the sample is known to be 2.2

    (b) The average household size in the sample is estimated to be 2.2

    (c) The average household size in the city is known to be 2.2

    (d) The average household size in the city is estimated to be 2.2

17. A survey company took a simple random sample of 275 rental flats out of all the rental flats in Bengaluru. The average monthly rent of the sampled units was ₹15000 and the SD was ₹9000. There were 964 people living in the sampled flats, and there were 120 children among the these 964 people. Pick the correct statements:

    (a) Average monthly rent of flats in Bengaluru is estimated to be ₹15000

    (b) This distribution will not be a normal distribution

    (c) 964 people is a good simple random sample to calculate the percentage of children living in the flats in Bengaluru

    (d) 964 people is not a good simple random sample to calculate the percentage of children living in the flats in Bengaluru

18. A list of numbers has been converted to standard units. See the table given below. Two of the cells in the table are unknown. Find it!

value	standard unit
54	2.8
A	-1.2
42	0.4
49	B

    (a) A=34 and B=1.8

    (b) A=40 and B=1.8

    (c) A=34 and B=1.5

    (d) A=40 and B=1.5

19. Piechart is mostly meant for

    (a) Ordinal variable

    (b) Nominal variable

    (c) Continuous variable

    (d) Discrete variable

20. At an Institute 70 students are raising money for Kerala flood relief fund. The average amount raised is ₹ 100. What is the largest possible number of children who could have raised ₹ 1000 or more?

    (a) 6

    (b) 7

    (c) 8

    (d) 9

Answer Key for Exam B

1. For the income data distribution given below, find the *inter-quartile range*, that is, the distance between the 75th percentile and the 25th percentile. In the calculation if you get decimals please round it off to two decimal places.

0 – \$10000	20%
\$10001 – \$25000	28%
\$25001 – \$50000	27%
\$50001 – \$100000	18%
\$100000 – \$150000	7%

   (a) \$37230
   (b) \$37320
   (c) \$37310
   (d) \$37240

2. Marks in a test are between 0 and 100. The average is 85 and the SD is 12. Based on this data, can we use the normal approximation to estimate percentiles of the distribution? If not, what shall we use to get a good estimation?

   (a) Normal approximation is enough
   (b) Normal approximation cannot be used we should use Markov's inequality
   (c) Normal approximation cannot be used we should use Chebechev's inequality
   (d) Not possible to decide we need more data

3. A simple random sample of 1900 households (How many people are there in a house) is taken in Jaipur. The average household size in the sample is 2.2 people, with an SD of 2 people. Now how about this statement: *Approximately 95% of the households had sizes in the range 2.07 to 2.33 people*

   (a) True statement
   (b) False statement
   (c) Valid statement but need to check the truthfulness with more data
   (d) Invalid statement

4. An Institute records the distance traveled from home to work of all of its 80 employees. By mistake, the largest commute distance is recorded as 65kms instead of 6.5kms. By how many miles does this mistake change the median of the commute distances? And how does the mean change?

   (a) Median by 16.5kms and Mean by 0.73kms
   (b) Median by 16.5kms and Mean by 0.72kms
   (c) Median by 0kms and Mean by 0.73kms
   (d) Median by 0kms and Mean by 0.72kms

For the following four MCQs use the below information:

We are going to test whether or not a random number generator is producing the digit 0 *in the correct proportion, i.e.,* 1/10. *For this we run the generator 5,000 times. We assume that the runs are mutually independent and that each has the same probability* p *of producing* 0.

We would like to construct a test that has a significance level of approximately 1%. Towards this answer the following MCQs.

5. The distribution of 0 follows what kind of a distribution (precisely)?

   (a) Normal
   (b) Poisson
   (c) Bernoulli
   [(d)] Binomial

6. What is the mean of the above distribution?

   [(a)] 500
   (b) 50
   (c) 5000
   (d) None of the above

7. What is the SD of the above distribution?

   (a) 21.12
   [(b)] 21.21
   (c) 22.36
   (d) 22.63

8. In order to test the hypothesis which of the following will be the most appropriate:

   [(a)] $H_0 : p = 0.1$ and $H_A : p \neq 0.1$
   (b) $H_0 : p = 0.1$ and $H_A : p \leq 0.1$
   (c) $H_0 : p = 0.1$ and $H_A : p \geq 0.1$
   (d) $H_0 : p \leq 0.1$ and $H_A : p \geq 0.1$

9. Here is a stem and leaf plot of the heights (in centimeters) of a group of men.
   16|56778
   17|0012234444
   17|55666788889
   18|0111234
   18|56
   How many men are 178cms tall and what is the range of the heights (i.e., difference between the smallest and the tallest height)?

   (a) 3 and 20cms
   (b) 3 and 21cms
   [(c)] 4 and 21cms
   (d) 4 and 20cms

10. For the income data distribution given below, find the *40th percentile* in thousands of dollars. In the calculation if you get decimals please round it off to two decimal places.

0 – \$10000	20%
\$10001 – \$25000	28%
\$25001 – \$50000	27%
\$50001 – \$100000	18%
\$100000 – \$150000	7%

[(a)] \$20710
(b) \$20720
(c) \$20170
(d) \$20180

11. Pick the correct statement:

[(a)] Histogram is for quantitative data
[(b)] Bar Chart is for categorical data
(c) Histogram is for categorical data
(d) Bar Chart is for quantitative data

12. Given is a list of numbers that has an average of 50 and an SD of 10. Is there a way to transform the list somehow so that the list has the SD of 30 but leaving the average the same?

[(a)] Yes
(b) No
(c) Not possible with this information
(d) Possible only if all are positive integers

13. A simple random sample of 1900 households (How many people are there in a house) is taken in Jaipur. The average household size in the sample is 2.2 people, with an SD of 2 people. The distribution of household size is

(a) Perfectly Normal
(b) Approximately Normal
[(c)] Is not Normal at all
(d) We can't say. Need more data

14. Which of the following does not make sense at all:

(a) Mode of a list of nominal values
(b) Mean of a list of ordinal values
[(c)] Mean of a list of nominal values
[(d)] Median of a list of ordinal values

15. Given a mark list of students of IDS course that has an average of 50 and an SD of 10. What proportion of students would have got marks less than 90?

(a) 3/4
(b) 1/4
(c) 1/16
[(d)] 15/16

16. A simple random sample of 1900 households (How many people are there in a house) is taken in Jaipur. The average household size in the sample is 2.2 people, with an SD of 2 people. Pick the correct (more appropriate) statement(s):

[(a)] The average household size in the sample is known to be 2.2

(b) The average household size in the sample is estimated to be 2.2

(c) The average household size in the city is known to be 2.2

[(d)] The average household size in the city is estimated to be 2.2

17. A survey company took a simple random sample of 275 rental flats out of all the rental flats in Bengaluru. The average monthly rent of the sampled units was ₹15000 and the SD was ₹9000. There were 964 people living in the sampled flats, and there were 120 children among the these 964 people. Pick the correct statements:

[(a)] Average monthly rent of flats in Bengaluru is estimated to be ₹15000

[(b)] This distribution will not be a normal distribution

(c) 964 people is a good simple random sample to calculate the percentage of children living in the flats in Bengaluru

[(d)] 964 people is not a good simple random sample to calculate the percentage of children living in the flats in Bengaluru

18. A list of numbers has been converted to standard units. See the table given below. Two of the cells in the table are unknown. Find it!

value	standard unit
54	2.8
A	-1.2
42	0.4
49	B

[(a)] A=34 and B=1.8

(b) A=40 and B=1.8

(c) A=34 and B=1.5

(d) A=40 and B=1.5

19. Piechart is mostly meant for

(a) Ordinal variable

[(b)] Nominal variable

(c) Continuous variable

(d) Discrete variable

20. At an Institute 70 students are raising money for Kerala flood relief fund. The average amount raised is ₹ 100. What is the largest possible number of children who could have raised ₹ 1000 or more?

(a) 6

[(b)] 7

(c) 8

(d) 9

The LNM Institute of Information Technology, Jaipur

CSE327 Introduction to Data Science

Quiz-I **Thursday, Sep 19, 2018**

Time: 45 Minutes Maximum Marks: 40

Name: _____ Roll No:_____

Read the Instructions Carefully:
All are objective questions with 4 choices. There may be more than one correct answers. Each question carry two marks. You need to select all the correct answers to get full marks. For example, if a question has 2 correct choices and you did only one correctly you will get 1 mark. If out of your choices even one choice is wrong then there won't be any marks given (even if there are correct choices that you have made). So be careful when you are finding and marking the correct choices. **Encircle all the correct answer/s. Any other marking will be considered as invalid**. Do not write anything in the answer sheet except the answer. No justification is required. **There is no negative marking for wrong answer/s.**

1. Marks in a test are between 0 and 100. The average is 85 and the SD is 12. Based on this data, can we use the normal approximation to estimate percentiles of the distribution? If not, what shall we use to get a good estimation?
   (a) Normal approximation is enough
   (b) Normal approximation cannot be used we should use Markov's inequality
   (c) Normal approximation cannot be used we should use Chebechev's inequality
   (d) Not possible to decide we need more data

2. At an Institute 70 students are raising money for Kerala flood relief fund. The average amount raised is ₹ 100. What is the largest possible number of children who could have raised ₹ 1000 or more?
   (a) 6
   (b) 7
   (c) 8
   (d) 9

3. A simple random sample of 1900 households (How many people are there in a house) is taken in Jaipur. The average household size in the sample is 2.2 people, with an SD of 2 people. The distribution of household size is
   (a) Perfectly Normal
   (b) Approximately Normal
   (c) Is not Normal at all
   (d) We can't say. Need more data

4. A simple random sample of 1900 households (How many people are there in a house) is taken in Jaipur. The average household size in the sample is 2.2 people, with an SD of 2 people. Now how about this statement: *Approximately 95% of the households had sizes in the range 2.07 to 2.33 people*
   (a) True statement
   (b) False statement
   (c) Valid statement but need to check the truthfulness with more data
   (d) Invalid statement

5. For the income data distribution given below, find the *inter-quartile range*, that is, the distance between the 75th percentile and the 25th percentile. In the calculation if you get decimals please round it off to two decimal places.

0 – $10000	20%
$10001 – $25000	28%
$25001 – $50000	27%
$50001 – $100000	18%
$100000 – $150000	7%

   (a) $37230
   (b) $37320
   (c) $37310
   (d) $37240

6. A survey company took a simple random sample of 275 rental flats out of all the rental flats in Bengaluru. The average monthly rent of the sampled units was ₹15000 and the SD was ₹9000. There were 964 people living in the sampled flats, and there were 120 children among the these 964 people. Pick the correct statements:
   (a) Average monthly rent of flats in Bengaluru is estimated to be ₹15000
   (b) This distribution will not be a normal distribution
   (c) 964 people is a good simple random sample to calculate the percentage of children living in the flats in Bengaluru
   (d) 964 people is not a good simple random sample to calculate the percentage of children living in the flats in Bengaluru

For the following four MCQs use the below information:

We are going to test whether or not a random number generator is producing the digit 0 *in the correct proportion, i.e.,* 1/10*. For this we run the generator 5,000 times. We assume that the runs are mutually independent and that each has the same probability p of producing* 0*.*

We would like to construct a test that has a significance level of approximately 1%. Towards this answer the following MCQs.

7. The distribution of 0 follows what kind of a distribution (precisely)?
   (a) Normal
   (b) Poisson
   (c) Bernoulli
   (d) Binomial

8. What is the mean of the above distribution?
   (a) 500
   (b) 50
   (c) 5000
   (d) None of the above

9. What is the SD of the above distribution?
   (a) 21.12
   (b) 21.21
   (c) 22.36
   (d) 22.63

10. In order to test the hypothesis which of the following will be the most appropriate:
   (a) $H_0 : p = 0.1$ and $H_A : p \neq 0.1$
   (b) $H_0 : p = 0.1$ and $H_A : p \leq 0.1$
   (c) $H_0 : p = 0.1$ and $H_A : p \geq 0.1$
   (d) $H_0 : p \leq 0.1$ and $H_A : p \geq 0.1$

11. Piechart is mostly meant for
   (a) Ordinal variable
   (b) Nominal variable
   (c) Continuous variable
   (d) Discrete variable

12. For the income data distribution given below, find the *40th percentile* in thousands of dollars. In the calculation if you get decimals please round it off to two decimal places.

0 – $10000	20%
$10001 – $25000	28%
$25001 – $50000	27%
$50001 – $100000	18%
$100000 – $150000	7%

(a) $20710

(b) $20720

(c) $20170

(d) $20180

13. An Institute records the distance traveled from home to work of all of its 80 employees. By mistake, the largest commute distance is recorded as 65kms instead of 6.5kms. By how many miles does this mistake change the median of the commute distances? And how does the mean change?

(a) Median by 16.5kms and Mean by 0.73kms

(b) Median by 16.5kms and Mean by 0.72kms

(c) Median by 0kms and Mean by 0.73kms

(d) Median by 0kms and Mean by 0.72kms

14. Given is a list of numbers that has an average of 50 and an SD of 10. Is there a way to transform the list somehow so that the list has the SD of 30 but leaving the average the same?

(a) Yes

(b) No

(c) Not possible with this information

(d) Possible only if all are positive integers

15. A simple random sample of 1900 households (How many people are there in a house) is taken in Jaipur. The average household size in the sample is 2.2 people, with an SD of 2 people. Pick the correct (more appropriate) statement(s):

(a) The average household size in the sample is known to be 2.2

(b) The average household size in the sample is estimated to be 2.2

(c) The average household size in the city is known to be 2.2

(d) The average household size in the city is estimated to be 2.2

16. Given a mark list of students of IDS course that has an average of 50 and an SD of 10. What proportion of students would have got marks less than 90?

(a) 3/4

(b) 1/4

(c) 1/16

(d) 15/16

17. A list of numbers has been converted to standard units. See the table given below. Two of the cells in the table are unknown. Find it!

value	standard unit
54	2.8
A	-1.2
42	0.4
49	B

(a) A=34 and B=1.8

(b) A=40 and B=1.8

(c) A=34 and B=1.5

(d) A=40 and B=1.5

18. Which of the following does not make sense at all:

(a) Mode of a list of nominal values

(b) Mean of a list of ordinal values

(c) Mean of a list of nominal values

(d) Median of a list of ordinal values

19. Here is a stem and leaf plot of the heights (in centimeters) of a group of men.

    16|56778
    17|0012234444
    17|55666788889
    18|0111234
    18|56

    How many men are 178cms tall and what is the range of the heights (i.e., difference between the smallest and the tallest height)?

    (a) 3 and 20cms

    (b) 3 and 21cms

    (c) 4 and 21cms

    (d) 4 and 20cms

20. Pick the correct statement:

    (a) Histogram is for quantitative data

    (b) Bar Chart is for categorical data

    (c) Histogram is for categorical data

    (d) Bar Chart is for quantitative data

Answer Key for Exam C

1. Marks in a test are between 0 and 100. The average is 85 and the SD is 12. Based on this data, can we use the normal approximation to estimate percentiles of the distribution? If not, what shall we use to get a good estimation?

   (a) Normal approximation is enough

   (b) Normal approximation cannot be used we should use Markov's inequality

   [(c)] Normal approximation cannot be used we should use Chebechev's inequality

   (d) Not possible to decide we need more data

2. At an Institute 70 students are raising money for Kerala flood relief fund. The average amount raised is ₹ 100. What is the largest possible number of children who could have raised ₹ 1000 or more?

   (a) 6

   [(b)] 7

   (c) 8

   (d) 9

3. A simple random sample of 1900 households (How many people are there in a house) is taken in Jaipur. The average household size in the sample is 2.2 people, with an SD of 2 people. The distribution of household size is

   (a) Perfectly Normal

   (b) Approximately Normal

   [(c)] Is not Normal at all

   (d) We can't say. Need more data

4. A simple random sample of 1900 households (How many people are there in a house) is taken in Jaipur. The average household size in the sample is 2.2 people, with an SD of 2 people. Now how about this statement: *Approximately 95% of the households had sizes in the range 2.07 to 2.33 people*

   (a) True statement

   (b) False statement

   (c) Valid statement but need to check the truthfulness with more data

   [(d)] Invalid statement

5. For the income data distribution given below, find the *inter-quartile range*, that is, the distance between the 75th percentile and the 25th percentile. In the calculation if you get decimals please round it off to two decimal places.

0 – $10000	20%
$10001 – $25000	28%
$25001 – $50000	27%
$50001 – $100000	18%
$100000 – $150000	7%

(a) $37230
[(b)] $37320
(c) $37310
(d) $37240

6. A survey company took a simple random sample of 275 rental flats out of all the rental flats in Bengaluru. The average monthly rent of the sampled units was ₹15000 and the SD was ₹9000. There were 964 people living in the sampled flats, and there were 120 children among the these 964 people. Pick the correct statements:

[(a)] Average monthly rent of flats in Bengaluru is estimated to be ₹15000
[(b)] This distribution will not be a normal distribution
(c) 964 people is a good simple random sample to calculate the percentage of children living in the flats in Bengaluru
[(d)] 964 people is not a good simple random sample to calculate the percentage of children living in the flats in Bengaluru

For the following four MCQs use the below information:

We are going to test whether or not a random number generator is producing the digit 0 *in the correct proportion, i.e.,* 1/10*. For this we run the generator 5,000 times. We assume that the runs are mutually independent and that each has the same probability p of producing* 0*.*

We would like to construct a test that has a significance level of approximately 1%. Towards this answer the following MCQs.

7. The distribution of 0 follows what kind of a distribution (precisely)?

(a) Normal
(b) Poisson
(c) Bernoulli
[(d)] Binomial

8. What is the mean of the above distribution?

[(a)] 500
(b) 50
(c) 5000
(d) None of the above

9. What is the SD of the above distribution?

(a) 21.12
[(b)] 21.21
(c) 22.36
(d) 22.63

10. In order to test the hypothesis which of the following will be the most appropriate:

    [(a)] $H_0 : p = 0.1$ and $H_A : p \neq 0.1$
    (b) $H_0 : p = 0.1$ and $H_A : p \leq 0.1$
    (c) $H_0 : p = 0.1$ and $H_A : p \geq 0.1$
    (d) $H_0 : p \leq 0.1$ and $H_A : p \geq 0.1$

11. Piechart is mostly meant for

    (a) Ordinal variable
    [(b)] Nominal variable
    (c) Continuous variable
    (d) Discrete variable

12. For the income data distribution given below, find the *40th percentile* in thousands of dollars. In the calculation if you get decimals please round it off to two decimal places.

0 – \$10000	20%
\$10001 – \$25000	28%
\$25001 – \$50000	27%
\$50001 – \$100000	18%
\$100000 – \$150000	7%

    [(a)] \$20710
    (b) \$20720
    (c) \$20170
    (d) \$20180

13. An Institute records the distance traveled from home to work of all of its 80 employees. By mistake, the largest commute distance is recorded as 65kms instead of 6.5kms. By how many miles does this mistake change the median of the commute distances? And how does the mean change?

    (a) Median by 16.5kms and Mean by 0.73kms
    (b) Median by 16.5kms and Mean by 0.72kms
    [(c)] Median by 0kms and Mean by 0.73kms
    (d) Median by 0kms and Mean by 0.72kms

14. Given is a list of numbers that has an average of 50 and an SD of 10. Is there a way to transform the list somehow so that the list has the SD of 30 but leaving the average the same?

    [(a)] Yes
    (b) No
    (c) Not possible with this information
    (d) Possible only if all are positive integers

15. A simple random sample of 1900 households (How many people are there in a house) is taken in Jaipur. The average household size in the sample is 2.2 people, with an SD of 2 people. Pick the correct (more appropriate) statement(s):

[(a)] The average household size in the sample is known to be 2.2

(b) The average household size in the sample is estimated to be 2.2

(c) The average household size in the city is known to be 2.2

[(d)] The average household size in the city is estimated to be 2.2

16. Given a mark list of students of IDS course that has an average of 50 and an SD of 10. What proportion of students would have got marks less than 90?

(a) 3/4

(b) 1/4

(c) 1/16

[(d)] 15/16

17. A list of numbers has been converted to standard units. See the table given below. Two of the cells in the table are unknown. Find it!

value	standard unit
54	2.8
A	-1.2
42	0.4
49	B

[(a)] A=34 and B=1.8

(b) A=40 and B=1.8

(c) A=34 and B=1.5

(d) A=40 and B=1.5

18. Which of the following does not make sense at all:

(a) Mode of a list of nominal values

(b) Mean of a list of ordinal values

[(c)] Mean of a list of nominal values

[(d)] Median of a list of ordinal values

19. Here is a stem and leaf plot of the heights (in centimeters) of a group of men.
16|56778
17|0012234444
17|55666788889
18|0111234
18|56
How many men are 178cms tall and what is the range of the heights (i.e., difference between the smallest and the tallest height)?

(a) 3 and 20cms

(b) 3 and 21cms

[(c)] 4 and 21cms

(d) 4 and 20cms

20. Pick the correct statement:

[(a)]	Histogram is for quantitative data
[(b)]	Bar Chart is for categorical data
(c)	Histogram is for categorical data
(d)	Bar Chart is for quantitative data