University of Dhaka

Department of Computer Science and Engineering

In-Course Examination

3rd Year 2nd Semester, Session: 2017-2018

STAT – 3205, Introduction to Probability and Statistics

Total Marks: 30 Time: 1 Hour 30 Minutes

(Answer any 5 of the following Questions)

1. a) Define 'variable', 'experimental unit' and 'measurement' with example. 2

b) Fifty people are grouped into four categories— A, B, C, and D—and the number of people who fall into each category is shown in the table:

Category	Frequency
A	11
B	14
C	20
D	5

i) What is the experimental unit? 0.5

ii) What is the variable being measured? Is it qualitative or quantitative? 1

iii) Construct a bar chart to describe the data. 1.5

iv) Does the shape of the bar chart in iii) change depending on the order of presentation of the four categories? Is the order of presentation important? 1

2. Suppose 30 students have achieved following scores in STAT-3205 course:

56 89 51 79 58 62 52 88 52 68 69 75 77 72 61
55 87 53 65 61 93 54 76 80 91 59 66 78 71 67

a) Use the range approximation to approximate the standard deviation of these 30 measurements. 1

b) Construct a relative frequency histogram to describe the data. 2

c) Construct a stem and leaf plot. 2

d) Comment on the shape of data distribution based on b) and c). 1

3. Given the following data set: 19, 12, 16, 0, 14, 8, 6, 7, 12, 13, 10, 19, 7, 5, 8

a) Find the five-number summary and the IQR. 2

b) Construct a box plot for these data and identify any outliers 2

c) Calculate the z-score for the smallest and largest observations. Is either of these observations unusually large or unusually small? 2

4. The table below shows the prices of nine smartphones along with their overall score (on a scale of 0–100) in a consumer rating survey

Brand and Model	Price($)	Overall Score
Motorola Droid X	200	75
Motorola Droid	150	73
HTC Droid	200	73
LG Ally	50	72
Samsung Omnia II	50	71
HTC Imagio	100	70
Motorola Devour	80	70
Blackberry Storm2 9550	150	70
Palm Pre Plus	50	66

a) Plot the nine data points using a scatterplot. 1

b) Calculate r, the correlation coefficient between price and overall score. 2

c) Comment on form, direction, and strength of the relationship between price and overall score. 1

d) Find the regression line for predicting the overall score of a smartphone based on its price. 2

5. a) Explain the difference between 'mutually exclusive' and 'independent' events. 3

b) A survey of people in a given region showed that 20% were smokers. The probability of death due to lung cancer, given that a person smoked, was roughly 10 times the probability of death due to lung cancer, given that a person did not smoke. If the probability of death due to lung cancer in the region is .006, what is the probability of death due to lung cancer given that a person is a smoker? 3

6. a) State and prove the law of total probability. 2

b) Suppose that, in a particular city, airport A, B and C handle 50%, 30% and 20% of all airline traffic, respectively. The detection rates for weapons at the three airports are .9, .8, and .85, respectively. If a passenger at one of the airports is found to be carrying a weapon through the boarding gate, what is the probability that the passenger is using airport B? 2

c) Two fair dice are tossed. What is the probability that the sum of the number of dots shown on the upper faces is less than 10? What is the probability that the number of one upper face is double than other upper face? 2

University of Dhaka
Department of Computer Science and Engineering
3rd Year 2nd Semester B.Sc. Final Examination, 2018
STAT-3205: Introduction to Probability and Statistics

Total Marks: 60 **Time: 3 Hours**

(Answer any 4 (Four) of the following Questions)

1. a) Six vehicles are selected from the vehicles that are issued campus parking permits, and the following data are recorded: [3]

Vehicle	Type	Make	Carpool?	One-way Commute Distance (miles)	Age of Vehicle (years)
1	Car	Honda	No	23.6	6
2	Car	Toyota	No	17.2	3
3	Truck	Toyota	No	10.1	4
4	Van	Dodge	Yes	31.7	2
5	Motor-cycle	Harley-Davidson	No	25.5	1
6	Car	Chevrolet	No	5.4	9

(i) What are the experimental units?
(ii) What are the variables being measured? What types of variables are they?
(iii) Is this univariate, bivariate, or multivariate data?

b) A discrete variable can take on only the values 0, 1, or 2. A set of 20 measurements on this variable is shown here: [2]

1	2	1	0	2
2	1	1	0	0
2	2	1	1	0
0	1	2	1	1

(i) Draw a dotplot to describe the data.
(ii) Comment on the shape of the distribution. Do you see any outliers?

c) The ages (in months) at which 50 children were first enrolled in a preschool are listed below. [5]

38	40	30	35	39	40	48	36	31	36
47	35	34	43	41	36	41	43	48	40
32	34	41	30	46	35	40	30	46	37
55	39	33	32	32	45	42	41	36	50
42	50	37	39	33	45	39	46	36	31

(i) Construct a stem and leaf display for the data.
(ii) Construct a relative frequency histogram for these data. Start the lower boundary of the first class at 30 and use a class width of 5 months.
(iii) Compare the graphs in parts i) and ii). Are there any significant differences that would cause you to choose one as the better method for displaying the data?

d) The length of time (in months) between the onset of a particular illness and its recurrence was recorded for $n = 50$ patients: [5]

2.1	4.4	2.7	32.3	2.5	9.8	2.8	6.6	2.9	1.6
14.7	8.6	18.7	7.4	8.2	19.2	5.9	4.3	3.3	1.2
4.1	18.4	1.2	6.1	13.5	2.4	2.1	8.3	2.8	1.3
14.1	1.0	2.5	2.4	18.0	8.7	24.0	1.4	8.2	5.8
1.6	3.5	11.4	18.0	26.7	3.7	12.6	23.1	5.8	4

Calculate the sample mean $\bar{x}$ as an estimate of μ, the mean length of time between the onset of a particular illness and its recurrence. Calculate s for the data. Construct the intervals $\bar{x} \pm s$, $\bar{x} \pm 2s$ and $\bar{x} \pm 3s$. Calculate the percentage of squares falling into each of the three intervals, and compare with the corresponding percentages given by the empirical rule and Tchebysheff's theorem.

2. a) Suppose that it is known that the number of bugs produced in a 2000 line code is a random variable with mean 50. [5]
(i) What can be said about the probability that the number of bugs will be more than 75?
(ii) If the variance of the number of bugs is known to equal 25, then what is the probability that the number of bugs in this 2000 line code will be between 40 and 60?

b) You are given $n = 10$ measurements: 3, 1, 5, 6, 4, 4, 3, 5, 6, 8.

(i) Find the mode and comment on modality.
(ii) Calculate the sample variance and standard deviation.
(iii) Compare the range and the standard deviation. The range is approximately how many standard deviations?

[6]

c) Given the following data set:
25, 22, 26, 23, 27, 26, 28, 12, 18, 25, 10, 30, 8, 11

(i) Find the five-number summary and the IQR.
(ii) Calculate the z-score for the smallest and largest observations. Is either of these observations unusually large or unusually small?
(iii) Construct a box plot for these data and identify any outliers

3. a) A group of items are categorized according to a certain attribute—X, Y, Z—and according to the state in which they are produced: [4]

	X	Y	Z
New York	20	5	5
California	10	10	5

(i) Graph these data using all possible methods to compare the numbers of items of each type made in California and New York.
(ii) Which of the charts best depicts the difference or similarity the numbers of items of each type made in California and New York?

b) A set of bivariate data consists of these measurements on two variables, x and y: [6]
(3, 6) (5, 8) (2, 6) (1, 4) (4, 7) (4, 6)

(i) Draw a scatterplot to describe the data.
(ii) Calculate the correlation coefficient, r. Comment on relationship between the two variables.
(iii) Calculate the equation of the regression line using the computing formulas.

c) What are the differences between 'mutually exclusive' and 'independent' events? [2]

d) A sample space S consists of five simple events with these probabilities:
$P(E_1) = P(E_2) = .15$, $P(E_3) = .4$ and $P(E_4) = 2P(E_5)$ [3]

(i) Find the probabilities for simple events E_4 and E_5.
(ii) Find the probabilities for these two events: $A = \{E_1, E_3, E_4\}$, $B = \{E_2, E_3\}$
(iii) List the simple events that are either in event A or event B or both.

4. a) A survey of people in a given region showed that 20% were smokers. The probability of death due to lung cancer, given that a person smoked, was roughly 10 times the probability of death due to lung cancer, given that a person did not smoke. If the probability of death due to lung cancer in the region is .006, what is the probability of death due to lung cancer given that a person is a smoker? [2]

b) A population can be divided into two subgroups that occur with probabilities 60% and 40%, respectively. An event A occurs 30% of the time in the first subgroup and 50% of the time in the second subgroup. What is the unconditional probability of the event A, regardless of which subgroup it comes from? [3]

c) While designing Moghbazar flyover a civil engineer was studying a right-turn lane that is long enough to hold seven cars. Let X be the number of cars in the lane at the end of a randomly chosen red light. The engineer believes that the probability that $X = x$ is proportional to $(x+1)(8-x)$ for $x = 0,...,7$ (the possible values of X). [6]

(i) Find the p.f. of X.
(ii) Find the probability that X will be at least 5.

d) Past experience has shown that, on the average, only 1 in 10 wells drilled hits oil. Let x be the number of drillings until the first success (oil is struck). Assume that the drillings represent independent events. [4]

(i) Find $p(1)$, $p(2)$, and $p(3)$.
(ii) Give a formula for $p(x)$.

Mention the characteristics of a binomial experiment. [3]

b) Over a long period of time, it has been observed that a professional basketball player can make a free throw on a given trial with probability equal to .8. Suppose he shoots four free throws. [3]

(i) What is the probability that he will make exactly two free throws?
(ii) What is the probability that he will make at least one free throw?

c) Let x be a hypergeometric random variable with $N = 15$, $n = 3$, and $M = 4$. [4]

(i) Calculate $p(1)$ and $p(3)$
(ii) Construct the probability histogram for x.
(iii) calculate $\mu = E(x)$ and σ.

d) How continuous random variables differ from discrete random variables? [1]

e) Suppose that a random variable X has the uniform distribution on the interval [−2, 8]. Find the p.d.f. of X and the value of Pr(0 < X < 7). [4]

6. a) State and explain central limit theorem. [5]

b) A soda filling machine is supposed to fill cans of soda with 12 fluid ounces. Suppose that the fills are actually normally distributed with a mean of 12.1 oz and a standard deviation of .2 oz. What is the probability that the average fill for a 6-pack of soda is less than 12 oz? [Given P(z = 1.22) = 0.8888.] [3]

c) Suppose a random sample of $n = 25$ observations is selected from a population that is normally distributed with mean equal to 106 and standard deviation equal to 12. [4]
i) Give the mean and the standard deviation of the sampling distribution of the sample mean $\bar{x}$.
ii) Find the probability that $\bar{x}$ exceeds 110. [Given P(z = 1.6) = 0.9452, P(z = 1.66) = 0.9515, P(z = 1.67) = 0.9525]

d) When it is appropriate to use the normal approximation to binomial probabilities? [3]

0·0256