AL/2016/31/E-I

අධ්‍යයන පොදු සහතික පත්‍ර (උසස් පෙළ) විභාගය, 2016 අගෝස්තු
கல்விப் பொதுத் தராதரப் பத்திர (உயர் தர)ப் பரீட்சை, 2016 ஓகஸ்ற்
General Certificate of Education (Adv. Level) Examination, August 2016

வ්‍යාපාර සංඛ්‍යානය I வணிகப் புள்ளிவிவரவியல் I **Business Statistics I**	**31 E I**	පැය දෙකයි இரண்டு மணித்தியாலம் ***Two hours***

Instructions:

- *Answer **all** questions.*
- *Write your **Index Number** in the space provided in the answer sheet.*
- *Statistical tables will be provided. Calculators are **not allowed**.*
- *Instructions are given on the back of the answer sheet. Follow those carefully.*
- *In each of the questions **1** to **50**, pick one of the alternatives from (1), (2), (3), (4), (5) which is **correct** or **most appropriate** and **mark your response on the answer sheet with a cross** (×) on the number of the correct option in accordance with the instructions given at the back of the answer sheet.*

1. Which of the following statements is true?

(1) Secondary data can be used when the data meets the purpose and needs of the study.
(2) If the information about salary issues of teachers are needed from teachers, the best method that should be used is the interview method.
(3) Sampling errors are present even in complete enumeration.
(4) The number of questions in a questionnaire should be between 20 and 50.
(5) The individual observations can be identified in a grouped frequency distribution.

2. Which of the following statements is/are true?

A - Both the Lorenz curve and Ogive are based on cumulative frequencies.
B - The sum of the areas of all the rectangles of a histogram is same as the area under the frequency polygon.
C - Harmonic mean is the most suitable average when each value of a data set should be given an equal weight.

(1) A only (2) B only (3) A and B only (4) A and C only (5) All A, B and C

3. The following stem and leaf diagram gives test scores of 32 students in an examination.

Stem	Leaf
3	4, 8
4	2, 3, 5
5	1, 2, 4, 7, 8, 9
6	0, 1, 3, 5, 6, 7, 8, 9
7	0, 1, 2, 3, 4, 5, 6
8	1, 2, 4
9	1, 3, 7

The semi-interquartile range of the distribution is

(1) 9 (2) 10 (3) 14 (4) 20 (5) 22

4. In a journey of 120 km, a man travels first 60 km in a speed of 30 km per hour and the remaining 60 km in a speed of 60 km per hour. The average speed per hour for the journey is

(1) 20 km. (2) 40 km. (3) 42 km. (4) 45 km. (5) 60 km.

5. The mean score of 150 students in a class is 60. The mean score of boys in the class is 55 and that of the girls is 70. The number of boys and girls in the class are

(1) 100, 50 (2) 50, 100 (3) 90, 60 (4) 80, 70 (5) 105, 45 respectively.

[see page two

6. The export of a commodity increased by 30% in 2007, decreased by 20% in 2008 and then increased by 40% in 2009. The average rate of increase of export per annum in the period 2007 - 2009 is

(1) –30 (2) 30 (3) $(90)^{1/3}$

(4) $(24000)^{1/3}$ (5) $[(100+30)(100-20)(100+40)]^{1/3} - 100$

7. In a frequency distribution, if the mean is less than the mode and the median is greater than the mean, but less than the mode, then the shape of the distribution is

(1) symmetrical (2) positively skewed (3) negatively skewed
(4) leptokurtic (5) platykurtic

8. In case of a frequency distribution with open-ended classes, the measure of dispersion which can be easily used is

(1) range. (2) mean deviation. (3) standard deviation.
(4) quartile deviation. (5) coefficient of variation.

9. Which of the following statements is/are true about Bowley's coefficient of skewness?

A - When the extreme observations are present Bowley's coefficient of skewness is more appropriate.
B - Bowley's coefficient of skewness is based only on the central 50% of the data distribution.
C - When a frequency distribution has open-ended classes, Bowley's coefficient of skewness can not be used.

(1) A only (2) B only (3) A and B only (4) B and C only (5) All A, B and C

10. If the sum of squares of the deviations of 10 observations from mean 50 is 250, then the coefficient of variation is

(1) 1.2% (2) 5% (3) 10% (4) 12% (5) 25%

11. Which of the following statements is/are true about a box–and–whiskers diagram?

A - Box–and–whiskers diagrams can be used to compare variances of two or more distributions.
B - Box–and–whiskers diagram represents summary of five measures of a data set.
C - Box–and–whiskers diagram can be used to detect outliers.

(1) A only (2) A and B only (3) A and C only (4) B and C only (5) All A, B and C

12. Unequal class intervals are sometimes used in a frequency distribution,

(1) to make the class frequencies smaller.
(2) to make class limits and class boundaries equal.
(3) to reduce the number of classes with few frequencies.
(4) to calculate the summary measures more easily.
(5) to make the areas of rectangles of a histogram proportional to the class frequencies.

13. Which of the following statements is/are true about the simple linear regression model?

A - Units of the regression coefficient are same as the units of the dependent variable.
B - If the regression coefficient of Y on X is greater than one, the regression coefficient of X on Y must be less than one.
C - If the regression coefficient of Y on X positive, the correlation coefficient r is also positive.

(1) A only (2) B only (3) A and C only (4) B and C only (5) All A, B and C

14. Which of the following statements is **not true**?

(1) Correlation coefficient is independent of the units of measurements.
(2) In the correlation analysis, both variables are assumed to be random variables.
(3) If the correlation coefficient between X and Y is zero, we can conclude that X and Y are independent.
(4) Correlation coefficient between X and Y is only a measure of linear relationship.
(5) If the correlation coefficient between X and Y is r, then the correlation coefficient between X + 10 and Y + 15 is also r.

15. In a simple linear regression analysis, the coefficient of determination was found to be $r^2 = 0.95$. Which of the following statements is/are true?

A - Regression model will correctly predict 95% of the time.
B - About 95% of the total variation of Y can be explained by the regression model.
C - If the regression coefficient of Y on X is positive, X and Y have a strong positive correlation.

(1) A only (2) B only (3) A and B only (4) B and C only (5) All A, B and C

[see page three

16. Which of the following statements is/are true about the probability?

A - If all the possible outcomes of a random experiment are equally likely, the probability of an event can be calculated without conducting the experiment.

B - The probability of an impossible event is a negative value.

C - If in N repeated trials an event E happens M times, then the probability of occurring E is given by $\frac{M}{N}$.

(1) A only (2) B only (3) A and C only (4) B and C only (5) All A, B and C

17. If A and B are any two events with $P(A) = P_1$, $P(B) = P_2$, and $P(A \cap B) = P_3$, then the conditional probability of A **does not** occur given B **does not** occur is

(1) $1 - P_1 - P_2 + P_3$ (2) $1 - P_3$ (3) $\frac{1-P_1-P_2+P_3}{1-P_2}$

(4) $\frac{1-P_1-P_2+P_3}{1-P_1}$ (5) $1 - P_2$

18. If A and B are independent events with $P(A') = \frac{3}{4}, P(A \cap B) = \frac{1}{3}$, $P(B) = C$, then the value of C is

(1) $\frac{1}{12}$ (2) $\frac{1}{9}$ (3) $\frac{1}{4}$ (4) $\frac{1}{2}$ (5) $\frac{3}{4}$

19. If A and B are two events with $P(A \cup B) = \frac{5}{6}$, $P(A \cap B)' = \frac{2}{3}$ and $P(A') = \frac{1}{2}$ then,

(1) A and B are mutually exclusive. (2) A and B are dependent events.

(3) $P(A \mid B) = \frac{2}{3}$ (4) $P(B \mid A) = \frac{1}{2}$

(5) A and B are independent.

20. The monthly demand for a certain item has the following probability distribution.

Demand (x)	1	2	3	4	5	6
Probability $f(x)$	0.10	0.15	0.20	0.25	0.17	0.13

If the cost of producing x items is given by C = 5000 + 200x, the expected cost is

(1) 726 (2) 1300 (3) 5000 (4) 5200 (5) 5726

21. Which of the following statements is true?

(1) If X is a discrete random variable with probability function $f(x)$ for $x = 0, 1, ..., n$. P($X > K$) can be expressed as $\sum_{x=K}^{n} f(x)$.

(2) If X is a continuous random variable with probability density function $f(x)$ for $0 < x < 15$. P ($X = 10$) can be calculated by substituting $x = 10$ in $f(x)$.

(3) If the expectation of the random variable X is negative, the variance of X is also negative.

(4) The expectation of the random variable X is same as the value of X which occurs with maximum probability.

(5) If $Var(X) = 1$, then $Var(2X \pm 3) = 4$.

22. For which of the following binomial distributions, the normal distribution provides the best approximation?

(1) $n = 100$ and $p = 0.02$ (2) $n = 100$ and $p = 0.05$

(3) $n = 100$ and $p = 0.50$ (4) $n = 100$ and $p = 0.95$

(5) $n = 100$ and $p = 0.98$

23. In a certain town, 50% of households is known to use a particular brand of soap. In a households survey, 1024 investigators are employed and each investigator takes a sample of 10 households. How many investigators are likely to report two houses which use this soap?

(1) 20 (2) 45 (3) 512 (4) 5100 (5) 5120

24. Suppose that flaws in certain cloth occur at random with an average of one flaw per 25 square meters. If a poisson distribution is assumed, what is the probability that a piece of cloth of 50 square meters will have at most one flaw?

(1) 0.135 (2) 0.271 (3) 0.406 (4) 0.736 (5) 0.982

25. If X is normally distributed with mean 3 and standard deviation 0.5 and $P(3 < X < C) = 0.4656$, then the value of C is

(1) -1.82 (2) 1.82 (3) 2.28 (4) 3.91 (5) 4.82

[see page four

26. Which of the following statements is true?

(1) The probability that a specific unit in a population of size N is included in a random sample of size n without replacement is $\frac{1}{N}$.

(2) In systematic sampling the term $\frac{n}{N}$ is called sampling interval.

(3) The finite population correction factor may be ignored when it is close to zero.

(4) Non-response errors may be reduced by increasing the sample size.

(5) Standard errors of estimators cannot be calculated in a non-probability sampling.

27. Which of the following statements is true?

(1) If the variation between stratas is large, the precision of stratified random sampling is also large.

(2) If intra-class correlation coefficient is close to one, cluster sampling is more efficient than simple random sampling.

(3) A quata sample is usually selected using a sampling frame.

(4) The main purpose of circular systematic sampling is to select a systematic sample when the sampling interval is an integer number.

(5) The field cost of cluster sampling is usually larger than that of simple random sampling.

28. Which of the following statements is true?

(1) The sampling distribution of the mean of a random sample from a normal distribution is normal only if the sample size is large.

(2) If the population mean μ is known $\frac{1}{n}\sum(x_i - \mu)^2$ is a biased estimator for the population variance σ^2.

(3) If the variance of an unbiased estimator $\hat{\theta}$ for θ approaches zero, when the sample size n approaches infinity, then $\hat{\theta}$ is called a consistent estimator for θ.

(4) If $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased estimators for θ, the efficiency of $\hat{\theta}_2$ with respect to $\hat{\theta}_1$ is measured by $\frac{Var(\hat{\theta}_2)}{Var(\hat{\theta}_1)} \times 100$.

(5) $\hat{\sigma} = \sqrt{\frac{1}{n-1}\sum(X_i - \bar{X})^2}$ is an unbiased estimator for σ.

29. Which of the following statements is/are true about the sampling distributions?

A - The t-distribution approaches the standard normal distribution as the number of degrees of freedom increases.

B - The shape of the F-distribution depends on the numerator degrees of freedom and the denominator degrees of freedom.

C - The Central Limit Theorem says that the sampling distribution of the sample mean is approximately normal for any sample size.

(1) A only (2) A and B only (3) A and C only (4) B and C only (5) All A, B and C

30. If $\bar{X}$ and $\bar{Y}$ are sample means of sample size 25 each from N (2, 16) and N (1, 9) respectively. $P(\bar{X} > \bar{Y})$ is

(1) 0.1587 (2) 0.3174 (3) 0.3413 (4) 0.6826 (5) 0.8413

31. If $\bar{X}$ is the mean of a random sample of size n from $N(\mu, 100)$, find the value of n such that $P(-5 < \bar{X} - \mu < 5) = 0.9544$.

(1) 4 (2) 8 (3) 15 (4) 16 (5) 18

32. Which of the following statements is/are true about the confidence intervals?

A - The width of a confidence interval increases as confidence level decreases.

B - Confidence intervals can be used to test some hypothesis.

C - For small samples, the width of the confidence intervals based on t-distribution is larger than the width of the confidence intervals based on Z-distribution.

(1) A only (2) B only (3) A and B only (4) B and C only (5) All A, B and C

33. The mean of a random sample of size 16 from $N(\mu, 100)$ distribution is observed as $\bar{x} = 114.5$. For testing H_0: $\mu = 112$ against H_1: $\mu \neq 112$, the P-value is

(1) 0.1587 (2) 0.1706 (3) 0.3413 (4) 0.6286 (5) 0.6826

[see page five

34. Which of the following statements is/are true about hypothesis testing?

A - If σ^2 of a normal population is unknown the hypothesis H_0: $\mu = \mu_0$ is a composite hypothesis.

B - If σ^2 of a normal population is unknown $Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$ is a test statistic.

C - If the probability of occurring type II error is β the power of the test is $1 - \beta$.

(1) A only (2) C only (3) A and C only (4) B and C only (5) All A, B and C

35. Which of the following statements is true?

(1) A hypothesis test with 10% significance level is better than a hypothesis test with 5% significance level.
(2) In a hypothesis test at 5% significance level, H_0 is not rejected if P-value < 0.05.
(3) The decision to use one-tailed or two-tailed test depends on the form of the null hypothesis.
(4) In a hypothesis test both type I error and type II error can be reduced only by increasing the sample size.
(5) The value obtained by substituting sample data to test statistic is called the critical value.

36. Let $\bar{X}$ be the mean of a random sample of size 25 from $N(\mu, 100)$ distribution. If the critical region for testing H_0: $\mu = 60$ against H_1: $\mu > 60$ is given by $\bar{X} > 63$, the probability of type I error is

(1) 0.0668. (2) 0.1336 (3) 0.2266 (4) 0.4332 (5) 0.5668

37. Suppose the proportions of male and female students who buy a certain magazine are π_1 and π_2 respectively. If 20 students of a random sample of 50 male students and 30 students of a random sample of 50 female students buy this magazine, the value of the test statistic for testing H_0: $\pi_1 = \pi_2$ is

(1) –2.11 (2) –2 (3) 2 (4) 2.11 (5) 2.34

38. A poisson distribution was fitted to a observed frequency distribution with ten classes, 0, 1, 2, ..., 9. The expected frequencies of last two classes are less than 5. The critical region for testing the goodness of fit at 5% level is given by

(1) $\chi^2 > 14.1$. (2) $\chi^2 > 15.5$. (3) $\chi^2 > 16.0$ (4) $\chi^2 > 16.9$ (5) $\chi^2 > 17.5$.

39. An analysis of variance table constructed to compare means of three normal populations with equal variances gives sums of square for between populations as 70 and sums of squares for errors as 36 with 12 degrees of freedom. The F-value in the analysis of variance table is

(1) 1.94 (2) 2.83 (3) 2.91 (4) 7.78 (5) 11.6

40. The estimated trend equation for annual sales of a commodity with 1st July 2014 as the origin is $\hat{Y} = 125 + 30.8X$ (unit X is one year). The trend equation that gives monthly trend values with 15th January 2015 as the origin is

(1) $\hat{Y} = \frac{125}{12} + \frac{30.8}{12}(X + 6.5)$ (2) $\hat{Y} = \frac{125}{12} + \frac{30.8}{144}(X + 7.5)$

(3) $\hat{Y} = \frac{125}{144} + \frac{30.8}{144}(X + 6.5)$ (4) $\hat{Y} = \frac{125}{12} + \frac{30.8}{144}(X + 7)$

(5) $\hat{Y} = \frac{125}{12} + \frac{30.8}{144}(X + 6.5)$

41. If the original time series value (Y) for each quater is expressed as the percentage of corresponding trend value and then take the average of the values for each quater over successive years, we get

(1) Deseasonalized data (2) Seasonal index
(3) Cyclical component (4) Irregular component
(5) Residual average

42. Under the multiplicative time series model, the deseasonalized data are obtained by

(1) subtracting the seasonal component (S) from the original data (Y).
(2) adding the seasonal component (S) to the original data (Y).
(3) multiplying the original data (Y) by the corresponding seasonal index (S).
(4) dividing the original data (Y) by the corresponding seasonal index (S).
(5) subtracting the trend component (T) from the original data (Y).

[see page six

43. Which of the following indices that does not reflect changes in buying patterns over time and tend to over estimate price changes?

(1) Paasche's price index (2) Fisher's price index.
(3) Marchell-Edgeworth price index. (4) Laspeyre's price index.
(5) Typical year price index.

44. Which of the following statements is/are true about index numbers?

A - Fisher's index is considered as an ideal index because it satisfies both time reversal test and factor reversal test.
B - Simple aggregate price index is independent of units of measurements.
C - If the total expenditure in current year with base year quantities as weights is expressed as the percentage of total expenditure in base year period, it is called Laspeyre's consumer price index.

(1) A only (2) A and B only (3) A and C only (4) B and C only (5) All A, B and C

45. The following are the index numbers of a commodity with 2010 as the base year.

Year	:	2010	2011	2012	2013	2014
Index number	:	100	120	150	180	225

If the base year is shifted to 2012, the new index number for 2014 is

(1) 30 (2) 120 (3) 125 (4) 150 (5) 170

46. The price relatives and weights of three commodities are given in the following table.

Commodity	A	B	C
Price relative	115	125	120
Weight	W_1	W_2	$W_3 = W_1 + W_2$

If the sum of the weights is 20 and the weighted price index is 119%, then the values of the weights are

(1) $W_1 = 3$, $W_2 = 7$ and $W_3 = 10$ (2) $W_1 = 7$, $W_2 = 3$ and $W_3 = 10$
(3) $W_1 = 10$, $W_2 = 20$ and $W_3 = 30$ (4) $W_1 = 20$, $W_2 = 3$ and $W_3 = 23$
(5) $W_1 = 20$, $W_2 = 119$ and $W_3 = 139$

47. Which of the following statements is true?

(1) Variation in quality due to negligence of operators is an example for chance variations.
(2) Control charts are also used in product control.
(3) Points falling outside control limits indicate that assignable causes are present.
(4) If R-chart is in control but $\overline{X}$-chart shows lack of control, the process is in a state of statistical control.
(5) The 3σ control limits for a P-chart are based on a poisson distribution.

48. If the sample points of a control chart shows some trends within control limits, it is an indication of

(1) the process is completely in control.
(2) the presence of only chance variation.
(3) the process does not indicate assignable causes.
(4) the process needs some adjustments.
(5) the level of the production process is in a stable condition.

49. During an examination of 10 pieces of cloth, the number defects observed are given below.

3, 4, 5, 2, 7, 4, 6, 0, 2, 3

The upper control limit of the C-chart based on 3σ is given by

(1) 0.496 (2) 0.504 (3) 3.6 (4) 5.6921 (5) 9.2921

50. Which of the following statements is/are true about product control?

A - Operating characteristic curve of a sampling plan is a graphic representation of the probability of acceptance for variation in fraction defective in the lot.
B - Lot Tolerance Percentage Defective (LTPD) is the lot quality which is considered to be good by the consumer.
C - In a single sampling plan, the maximum allowable number of defectives in the sample is called acceptance number.

(1) A only (2) A and B only (3) A and C only (4) B and C only (5) All A, B and C

* * *

AL/2016/31/E-II

අධ්‍යයන පොදු සහතික පත්‍ර (උසස් පෙළ) විභාගය, 2016 අගෝස්තු
கல்விப் பொதுத் தராதரப் பத்திர (உயர் தர)ப் பரீட்சை, 2016 ஓகஸ்ற்
General Certificate of Education (Adv. Level) Examination, August 2016

ව්‍යාපාර සංඛ්‍යානය வணிகப் புள்ளிவிவரவியல் **Business Statistics**	**II** **II** **II**	**31 E II**	පැය තුනයි மூன்று மணித்தியாலம் ***Three hours***

Instructions:

* *Answer **five** questions selecting at least **two** questions from each part.*
* *Statistical tables and graph papers will be provided. Calculators are **not allowed**.*

Part I

1. (a) Explain the uses of the following statistical techniques in the field of business and industry.
(i) Statistical quality control
(ii) Time series analysis
(iii) Index numbers (06 *marks*)

(b) Explain the difference between sampled population and target population. Describe **four** advantages of a sample survey as compared with complete enumeration. (04 *marks*)

(c) Describe how would you construct the following charts using suitable examples.
(i) Z chart
(ii) Profile chart (04 *marks*)

(d) The following table gives the cumulative percentages of number of firms and net output.

No. of firms (%)	**Net output (%)**
40	08
82	37
88	48
95	61
97	74
100	100

Draw a Lorenz curve for these data and comment on the inequality of the distribution. (06 *marks*)

2. (a) What are properties of a good average? Examine these properties with reference to mean, median and mode. (05 *marks*)

(b) The age distribution of employees in a certain factory is given in the following table.

Age group	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64
No. of employees	60	70	60	50	40	30	30	25	20	15

(i) Calculate mean, median, mode and variance of the age distribution of employees in the factory.
(ii) Calculate Karl Pearson's coefficient of skewness and comment on the shape of the distribution. (10 *marks*)

(c) The mathematics marks of students in a certain class are given in the following cumulative distribution.

Marks (more than)	0	10	20	30	40	50	60	70	80	90
Number of students	80	76	74	70	60	46	37	10	05	02

Draw an Ogive for this distribution and find the percentage of students who pass the examination if the pass mark is 36. (05 *marks*)

[see page two

3. (a) Distinguish between consumer price index and whole sale price index. (03 *marks*)

(b) Using the following data, calculate

(i) Laspeyer's price index.

(ii) Paache's price index.

(iii) Fisher's price index.

Show that Fisher's price index satisfies both time reversal test and factor reversal test.

Commodity	Base year		Current year	
	Price (Rs.'00)	Quantity (kg)	Price (Rs.'00)	Quantity (kg)
A	2	7	6	6
B	3	6	2	3
C	4	5	8	5
D	5	4	2	4

(07 *marks*)

(c) Explain what do you mean by trend in time series analysis.

The annual production of a commodity (in '000) is given in the following table.

Year	2010	2011	2012	2013	2014	2015
Production ('000 units)	101	107	113	121	136	148

(i) Estimate the trend equation by using the method of least squares.

(ii) If the origin is shifted to year 2014, what is the new trend equation?

(iii) Forecast the production for the year 2017 by using new trend equation. (10 *marks*)

4. (a) The following table gives ages (X) of cars in years and annual maintenance cost (Y) (in thousand rupees).

Age of cars (X)	1	3	5	7	9
Maintenance cost (Y)	15	18	21	23	22

$\sum X = 25$, $\sum Y = 99$, $\sum X^2 = 165$, $\sum Y^2 = 2003$, $\sum XY = 533$

(i) Assuming a linear relationship, estimate the least squares regression equation for the above data.

(ii) Interpret the regression coefficient.

(iii) Estimate the expected maintenance cost, if the age of a car is 8 years. (10 *marks*)

(b) The following are the sample means and ranges for 10 random samples each of size 5. Calculate the control limits for mean chart and the range chart and state whether the process is in control or not.

Sample number	1	2	3	4	5	6	7	8	9	10
Mean	9.2	9.8	8.8	9.6	9.0	7.6	8.4	7.6	8.6	8.0
Range	5	2	6	3	5	2	6	2	5	7

(06 *marks*)

(c) Suppose that a large lot of goods was received a buyer. The acceptance sampling plan is to inspect a random sample of 100 goods and acceptance number is 2. Find the probability that the lot will be accepted, if it has 5% defectives. (04 *marks*)

[see page three

Part II

5. (a) Describe the classical approach and relative frequency approach of probability. Give **two** limitations for each. (04 *marks*)

(b) A salesman visits three houses to sell two products A and B. The probability that on any visit he makes a sale of product A is $\frac{2}{3}$ and the probability of he makes a sale of product B is $\frac{3}{4}$. If the probabilities of selling product A and B are independent, calculate the probability that the salesman will

(i) sell both products A and B at the visit of the first house.

(ii) sell one product at the visit of the first house.

(iii) make no sale of product B in three houses.

(iv) make at least one sale of product B in three houses. (08 *marks*)

(c) State the Bayes' theorem.

The germinate percentage of bean seeds from supplier A is 85% and the germinate percentage from supplier B is 75%. A seed packaging company purchases 40% of their been seeds from supplier A and 60% from supplier B and mixes these seeds together. If a randomly selected seed from the mixed seeds germinates, find the probability that the seed was purchased from supplier B. (08 *marks*)

6. (a) (i) Define the binomial distribution.

(ii) Describe the random experiment relevant to binomial distribution.

(iii) A multiple choice question paper consists of 10 questions with 5 answers to each question of which one is correct. A student answers each question by selecting one answer randomly from five answers. To get a distinction pass he must get at least 70% correct answers. What is the probability that the student will get a distinction pass? (06 *marks*)

(b) (i) Define the poisson distribution.

(ii) Apples are packed in boxes, each box containing 200. It is found that 0.5% of apples were rotten when boxes are opened. Using the poisson approximation, find the probabilities of 0 rotten apples, 1 rotten apple, 2 rotten apples in a box.

(iii) A buyer opens a box and if it contains more than one rotten apple, he opens another box and if it also contains more than one rotten apple he does not buy the boxes. What is the probability that he refuses to buy the boxes of apples? (06 *marks*)

(c) A student leaves home at 7.00 a.m. every morning in order to arrive at the school 8.00 a.m. He finds that over a long period he is late once in twenty times. Then he tries leaving home 6.55 a.m. and finds that he is late once in forty times. Assuming that the time of his journey has a normal distribution, before what time should he leave home in order not to be late more than once in 50 times? (08 *marks*)

7. (a) Describe the following methods of random sampling giving situations where each can be applied.

(i) Stratified random sampling (ii) Systematic sampling (iii) Cluster sampling (06 *marks*)

(b) In a population with $N = 6$, the values of y_i are 8, 2, 1, 10, 3, 6.

(i) Calculate the population mean $\overline{Y}$ and population variance S^2.

(ii) Calculate the sample mean $\overline{y}$ for all possible simple random samples of size 2 and obtain the sampling distribution of $\overline{y}$. Verify that the sample mean $\overline{y}$ is an unbiased estimator for population mean $\overline{Y}$. Find the variance of $\overline{y}$ using the sampling distribution of $\overline{y}$.

(iii) Verify that variance of $\overline{y}$ can be found without taking all possible samples using a formula since population variance S^2 is known. (10 *marks*)

(c) State the Central Limit Theorem. If a random sample of size 50 is taken from a poisson distribution with mean $\lambda = 2$, find the approximate probability that the sample mean will exceed 3. (04 *marks*)

[see page four

8. (a) (i) What do you mean by unbiasedness and efficiency of an estimator for a population parameter?

(ii) Let X_1, X_2, X_3 is a random sample from a population with mean μ and variance σ^2. If two estimators T_1 and T_2 for μ are defined as $T_1 = 2X_1 + 3X_2 - 4X_3$, $T_2 = \frac{1}{3}(X_1 + X_2 + X_3)$. Show that both are unbiased estimators for μ.

(iii) Which is the more efficient estimator? (06 *marks*)

(b) Samples of two types of electric bulbs were tested for length of life and the following data were observed.

Data observed	**Type I**	**Type II**
Number of bulbs used	08	07
Sample mean (hours)	1134	1024
Sample variance	1225	1600

Stating the assumption you make,

(i) Construct a 95% confidence interval for the difference in mean length of life.

(ii) Using the confidence intervals, test the hypothesis that the mean length of life of two types of bulbs are equal. (08 *marks*)

(c) The following table gives the classification of 100 students according to age and grade points.

Grade point	**Age (years)**			**Total**
	25 or below	**26 - 28**	**Over 28**	
Up to 3.0	06	09	05	20
3.1 to 3.5	18	14	08	40
3.6 to 4.0	11	12	17	40
Total	35	35	30	100

Test the hypothesis that age and grade points are independent at 5% significance level. (06 *marks*)

***