ශ්‍රී ලංකා විභාග දෙපාර්තමේන්තුව ශ්‍රී ලංකා විභාග දෙපාර්තමේන්තුව
இலங்கைப் பரீட்சைத் திணைக்களம்
Department of Examinations, Sri Lanka

අධ්‍යයන පොදු සහතික පත්‍ර (උසස් පෙළ) විභාගය, 2018 අගෝස්තු
கல்விப் பொதுத் தராதரப் பத்திர (உயர் தர)ப் பரீட்சை, 2018 ஓகஸ்ற்
General Certificate of Education (Adv. Level) Examination, August 2018

ව්‍යාපාර සංඛ්‍යානය I வணிகப் புள்ளிவிவரவியல் I **Business Statistics I**	**31 E I**

14.08.2018 / 1300 - 1500

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) Non-sampling error cannot occur in a complete census.
(2) In general target population is different from the sampled population.
(3) Pre-test is done after collecting data.
(4) Both sampling error and non-sampling error are included in standard error.
(5) Sampling frame is the list of sampling units in the selected sample.

2. Which of the following statement/s is/are true?

A - The area covered by the frequency polygon is same as the sum of areas of rectangles in the corresponding histogram.
B - The Lorenze curve is a graphical method of indicating whether a given quantity is equally distributed throughout the relevant population.
C - The mode can be easily obtained using the ogive of a distribution.

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

3. Which of the following statements is true?

(1) The width of a certain class interval can be obtained by subtracting the lower class limit from the upper class limit.
(2) If a constant A is added to each value of a data set then the variance of the data set will increase by A.
(3) A histogram cannot be constructed, if class widths are not equal.
(4) The class mark represents a class interval, if the data of the class interval distribute uniformly in the class interval.
(5) The class limit is also called the class boundary.

4. The most suitable chart to represent the total value with its components is

(1) Simple bar chart. (2) Multiple bar chart. (3) Pie chart.
(4) Pictograms. (5) Profile chart.

5. Which of the following statement/s is/are true?

A - Original data can be recovered from a stem-and-leaf diagram.
B - The shape of a distribution can be identified by looking at a stem-and-leaf diagram.
C - If the two boxes of a box-and-wishker diagram are equal then the distribution is exactly symmetric.

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

[see page two

6. Which of the following statement/s is/are true?

A - The median of a distribution is not affected by extream values.
B - Mode cannot be calculated for a distribution with unequal class intervals.
C - Geometric mean of a data set cannot be calculated when one value is negative.

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

7. In a certain factory a unit of work is completed by A in 3 minutes, by B in 5 minutes, by C in 6 minutes, by D in 10 minutes. The average rate of working time by these four persons in minutes is

(1) 3.20 (2) 5.00 (3) 5.47 (4) 5.50 (5) 6.00

8. In a distribution, the difference of first and third quartiles is 20 and their sum is 40. If the median of the distribution is 25, the value of Bowley's coefficient of skewness is

(1) –1.50 (2) –1.00 (3) – 0.50 (4) 0.50 (5) 0.75

9. Which of the following statement/s is/are true?

A - The kurtosis of a distribution is measured relative to the peakedness of a normal curve.
B - A more peaked curve relative to a normal curve is called mesokurtic.
C - Measures of kurtosis are not relevant for a skewed distribution.

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

10. If the mean of 10 values was 12 and the sum of squares was 1600, the coefficient of variation is

(1) 33.33% (2) 35.25% (3) 75.00% (4) 133.33% (5) 300.00%

11. Which of the following statement/s is/are true?

A - The arithmetic mean of a frequency distribution is a weighted average, the weight being the frequencies of classes.
B - If a distribution has the longer tail towards left, it is said to be a positively skewed distribution.
C - For a symmetric distribution Q_3 – Median = Median – Q_1.

(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. The geometric mean of four values was calculated as 16. Later it was found that the value 132 was wrongly recorded as 22. The corrected value of geometric mean is given by

(1) $(16)^{\frac{3}{2}}$ (2) $16\left(\frac{1}{6^{\frac{1}{4}}}\right)$ (3) $16\left(\frac{1}{6^{\frac{1}{2}}}\right)$ (4) $16(6^{\frac{1}{2}})$ (5) $16(6^{\frac{1}{4}})$

13. If the correlation coefficient between X and Y is r and $U=\frac{X}{h}$, $V=\frac{Y}{k}$, then the correlation coefficient between U and V is

(1) $\frac{r}{hk}$ (2) $\frac{r^2}{hk}$ (3) $\frac{r}{\sqrt{hk}}$ (4) r (5) hkr

14. Which of the following statement/s is/are true?

A - The regression line gives the means of Y for given values of X.
B - In the method of least squares, we assume that only the dependent variable is the variable which has error and independent variable has no error.
C - If in the regression of Y on X, the coefficient is b_{YX} and in the regression of X on Y the coefficient is b_{XY}, then $r^2 = b_{YX} \cdot b_{XY}$

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

[see page three

15. Which of the following statement/s is/are true?

A - If in a debate contest, the rank correlation coefficient between two judges is close to –1, it indicates that judges are strongly agree on the judgement.

B - Spearman's rank correlation coefficient is same as the product moment correlation coefficient between ranks.

C - If X and Y are continuous variables, the rank correlation coefficient cannot be calculated between observed values of X and Y.

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

16. Which of the following statement/s is/are true?

A - Under the classical approach the true probability of an event can be calculated without performing the experiment.

B - If the experimental conditions change when repeating the experiment, the more appropriate approach of probability is the relative frequency approach.

C - The subjective probability approach cannot be applied, if the outcomes of an experiment are not equally likely.

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

17. If A and B are two events with $P(A)=p_1$, $P(B)=p_2$ and $P(A\cap B)=p_3$, then the value of $P[A'\cap(A\cup B)]$ is

(1) $1-p_1-p_2+p_3$ (2) p_2+p_3 (3) $1-p_3$
(4) $p_1+p_2-p_3$ (5) p_2-p_3

18. If A and B are independent events with $P(A)<P(B)$, $P(A\cap B)=\frac{6}{25}$ and $P(A|B)+P(B|A)=1$, the value of $P(A)$ is,

(1) $\frac{1}{25}$ (2) $\frac{1}{5}$ (3) $\frac{6}{25}$ (4) $\frac{2}{5}$ (5) $\frac{3}{5}$

19. If A, B and C are three events, the probability that exactly one of them occur is given by

(1) $P(A\cup B\cup C)$ (2) $P(A\cap B\cap C)$
(3) $P(A'\cup B'\cup C')$ (4) $P(A\cap B'\cap C')+P(A'\cap B\cap C')+P(A'\cap B'\cap C)$
(5) $1-P(A\cup B\cup C)$

20. The random variable X has the following probability distribution.

x	–2	–1	0	1	2	3
$f(x)$	0.1	c	0.2	$2c$	0.3	c

The smallest value of x for which $P(X\le x)>0.5$ is

(1) –2 (2) –1 (3) 0 (4) 1 (5) 2

21. Which of the following statements is true?

(1) If Var(X) = 2, then Var($2X$+5) = 13.

(2) The expected value of the random variable X is same as the value of X occurred with the maximum probability.

(3) If X is a random variable with mean μ and variance σ^2, then $\text{Var}\left(\frac{X-\mu}{\sigma}\right)=1$.

(4) If X and Y are any two random variables, $E(XY)=E(X)\,E(Y)$.

(5) A mean of discrete random variable cannot have decimal values.

22. If for a binomial distribution $n=6$ and $9P(X=4)=P(X=2)$, then probability of success is

(1) $\frac{1}{9}$ (2) $\frac{1}{8}$ (3) $\frac{1}{4}$ (4) $\frac{1}{2}$ (5) $\frac{3}{4}$

[see page four

23. Which of the following statement/s is/are true?

A - The number of defective items in n randomly selected items with replacement from M items which consists of 'K' defective items has a binomial distribution.

B - The variance of the binomial distribution may exceed the mean of the binomial distribution.

C - For large n, the binomial distribution can be approximated by a poison distribution if the probability of the success is also large.

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

24. It has been found that 1% of the items produced by a certain machine are defective. The probability that in a random sample of 200 items, there are at most two defective items is approximately

(1) 0.3233 (2) 0.4060 (3) 0.6767 (4) 0.9814 (5) 0.9998

25. If X is normally distributed with mean 10 and $P(X<12) = 0.8413$ then the value of $P(9 \leq X \leq 11)$ is

(1) 0.1915 (2) 0.3413 (3) 0.3830 (4) 0.6826 (5) 0.9544

26. Which of the following statements is true?

(1) The accuracy of an estimate is measured by the standard error of that estimator.
(2) The variance of the sample mean in sampling with replacement is smaller than variance of the sample mean in sampling without replacement.
(3) The standard error of an estimator can be measured only in a probability sampling.
(4) The failure to interview the units in the selected sample is an example for a sampling error.
(5) The term $\frac{N}{n}$ is called the sampling fraction.

27. Which of the following statements is true?

(1) In stratified random sampling, the differences among strata are included in sampling error.
(2) A quota sample is selected using a sampling frame.
(3) The efficiency of the systematic sampling depends on the structure of the population.
(4) If the intra-class correlation is close to one, cluster sampling is more efficient than simple random sampling.
(5) In simple random sampling, the standard error of an estimator cannot be calculated using a single sample.

28. Which of the following statements is true?

(1) If μ is the unknown population mean, then $\frac{1}{n\text{-}1}\Sigma(x_i - \mu)^2$ is an unbiased estimator for σ^2.
(2) Since the sample variance S^2 is an unbiased estimator for σ^2, sample standard deviation S is also an unbiased estimator for σ.
(3) If both bias and variance of an estimator approaches zero when the sample size increases, it is a consistent estimator.
(4) If the value of an estimator is equal to the population parameter, it is an unbiased estimator.
(5) An estimator with minimum variance is called a sufficient estimator.

29. Which of the following statement/s is/are true?

A - According to the central limit theorem for sufficiently large n, the sampling distribution of the sample proportion has a normal distribution approximately.

B - For a given significance level, the table value taken from t-table is smaller than the value taken from Z-table.

C - F-distribution is used for comparing means of several normal populations with unequal variances.

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

[see page five

30. Which of the following formulas gives the standard error of the sample proportion p, from a finite population of size N?

(1) $\sqrt{\frac{N-n}{N}\frac{\pi(1-\pi)}{n}}$ (2) $\sqrt{\frac{N-n}{N-1}\frac{\pi(1-\pi)}{n}}$ (3) $\sqrt{\frac{N-1}{N-n}\frac{\pi(1-\pi)}{n}}$

(4) $\sqrt{\frac{\pi(1-\pi)}{n}}$ (5) $\sqrt{\frac{N-1}{N-n}\frac{\pi(1-\pi)}{n-1}}$

31. If the mean of a random sample of size 80 taken from a population with mean 128 and variance 20, the approximate probability that $\bar{X}$ lies between 127 and 129 is

(1) 0.2280 (2) 0.3413 (3) 0.4772 (4) 0.6826 (5) 0.9544

32. Which of the following statement/s is/are true?

A - The confidence limits for the population mean can be expressed as (point estimate) ± (table value) * (standard error).

B - Confidence intervals can be used for testing some hypothesis.

C - Confidence intervals with 99% level is wider than confidence intervals with 95% level.

(1) A only (2) A and B only (3) A and C only

(4) B and C only (5) All A, B and C

33. In testing the mean of a normal population as a two-tailed test, the $Z-$value was observed as $Z=1.4$. The $P-$value for this test is

(1) 0.0808 (2) 0.1616 (3) 0.4192 (4) 0.5808 (5) 0.8384

34. Which of the following statement/s is/are true?

A - When the significance level of a hypothesis test increases, the power of the test will decrease.

B - If the variance of a normal population is unknown, the hypothesis $H_0: \mu = 100$ is a simple hypothesis.

C - In hypothesis testing, both types of error can be reduced only by increasing the sample size.

(1) A only (2) B only (3) C only

(4) A and C only (5) All A, B and C

35. Which of the following statements is true?

(1) A test statistic cannot have parameters.

(2) The $P-$value for a test is calculated under the assumption that the alternative hypothesis is true.

(3) The sampling distribution of a test statistic is decided under the assumption that the alternative hypothesis is true.

(4) The observed value for a test statistic is called a critical value.

(5) The probability of accepting the correct null hypothesis is called the power of the test.

36. Let $\bar{X}$ be the mean of a random sample of size 100 from a distribution with $\sigma = 25$. If the critical region for testing $H_0: \mu = 50$ against $H_1: \mu = 60$ is given by $\bar{X} > 55$, the power of the test is

(1) 0.3413 (2) 0.3830 (3) 0.4772 (4) 0.6826 (5) 0.9772

37. Which of the following statement/s is/are true?

A - If $P-$value > 0.05, the null hypothesis should be rejected at 5% significance level.

B - The $P-$value of a test is calculated using the observed value for the test statistic and the relevant statistical table.

C - The confidence interval for parameters cannot be calculated using the sampling distribution of a test statistic.

(1) A only (2) A and B only (3) A and C only

(4) B and C only (5) All A, B and C

[see page six

ශ්‍රී ලංකා විභාග දෙපාර්තමේන්තුව ශ්‍රී ලංකා විභාග දෙපාර්තමේන්තුව
இலங்கைப் பரீட்சைத் திணைக்களம்
Department of Examinations, Sri Lanka

අධ්‍යයන පොදු සහතික පත්‍ර (උසස් පෙළ) විභාගය, 2018 අගෝස්තු
கல்விப் பொதுத் தராதரப் பத்திர (உயர் தர)ப் பரீட்சை, 2018 ஓகஸ்ற்
General Certificate of Education (Adv. Level) Examination, August 2018

ව්‍යාපාර සංඛ්‍යානය II வணிகப் புள்ளிவிவரவியல் II **Business Statistics II**	**31**	**E**	**II**

16.08.2018 / 08 30 - 11 40

පැය තුනයි
மூன்று மணித்தியாலம்
Three hours

අමතර කියවීම් කාලය - මිනිත්තු 10 යි
மேலதிக வாசிப்பு நேரம் - 10 நிமிடங்கள்
Additional Reading Time - 10 minutes

Use **additional reading time** to go through the question paper, select the questions and decide on the questions that you give priority in answering.

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) Describe **three** uses and **three** limitations of statistics. (03 *marks*)

(b) Describe what you mean by primary data and secondary data.
State **two** advantages and **two** disadvantages of each type of data. (04 *marks*)

(c) Explain the role of a pre-test and a pilot survey in conducting a sample survey. (04 *marks*)

(d) What are the points that should be taken into account when constructing a data table?

In the year 2014, out of a total 2000 workers of a factory, 1500 workers were permanent. The number of women workers was 300 out of which 200 were temporary. In the year 2017, the number of workers increased to 2800 of which 2000 were men. On the other hand the number of temporary workers fall down to 250 of which 150 were women. Present the above data in the form of an appropriate table. (04 *marks*)

(e) The age distribution of the workers of a certain factory is given below.

Age	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59
No. of workers	30	35	55	80	70	65	40	25

Draw 'the less than ogive' and find the median age of workers using it. (05 *marks*)

2. (a) Explain the importance of the following measures for identifying the shape of the distribution of a data set.

(i) Measures of Central Tendency
(ii) Measures of Variability
(iii) Measures of Skewness
(iv) Measures of Kurtosis (06 *marks*)

(b) The wages of 60 workers in a factory are given in the following distribution.

Wages (Rs.' 000)	20-24	25-29	30-34	35-39	40-44	45-49	50-54
No. of workers	03	10	20	15	05	04	03

(i) Calculate the mean, median, mode and standard deviation of the distribution.
(ii) Calculate Karl Pearson's coefficient of skewness and comment on the distribution. (08 *marks*)

[see page two

(c) Distinguish between absolute variation and relative variation.

In an examination, the mean mark of a group of 150 students for Mathematics was 78 and the standard deviation was 8. The mean mark of the group for statistics was 73 and the standard deviation was 7. For which subject would the

(i) absolute variation be higher value?

(ii) relative variation be higher value? (06 *marks*)

3. (a) Describe **three** types of index numbers which are commonly used.

What are the main problems involved in the construction of a cost of living index? Explain **three** main uses of a cost of living index. (06 *marks*)

(b) The following table provides the group index and the weight of various expenditure groups of a particular worker group for 2010 and 2015.

Expenditure Group	Group Index		Weight
	2010	2015	
Food	150	170	40
Fuel	20	30	10
Textiles	70	80	20
House Rent	30	40	10
Miscellaneous	40	50	20

What should be the rate of increase of wage in 2015 in order to maintain the standard of living of the worker group at the 2010 level? (04 *marks*)

(c) Describe the adjustments which are necessary in the time series raw data before making an analysis. (04 *marks*)

(d) The following table provides the annual crude birth rate of Sri Lanka from 2005 to 2017.

Year	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Crude Birth Rate	18.9	18.8	19.3	18.5	18.0	17.6	17.4	17.5	17.8	16.9	16.0	15.6	15.0

Predict the crude birth rate for the year 2018 by fitting a trend line by using the method of semi-average.

State the limitations of your prediction. (06 *marks*)

4. (a) Explain how you would use the method of least squares for fitting the regression line of Y on X.

The following table shows the amount of fertilizer used (X) and the yield of a certain crop (Y).

Fertilizer (grammes) (X)	0	1	2	3	4	5
Yield (kg) (Y)	15	20	30	40	50	60

(i) Construct a scatter diagram and comment on the relationship between X and Y.

(ii) Fit the regression line of Y on X by the method of least squares.

(iii) Calculate the coefficient of determination and interpret it.

(iv) Estimate the expected yield for X=7 and comment on the validity of the estimation. (10 *marks*)

(b) Explain the difference between the terms of each pair given below.

(i) P-chart and C-chart

(ii) Acceptable Quality Level (AQL) and Lot Tolerance Proportion Defective (LTPD) (04 *marks*)

[see page three

(c) Suppose a large lot is shipped to a certain company. The acceptance sampling plan is to accept the lot if the number of defectives of a sample of 100 is less than or equal to 2.

(i) If the defective percentage of the lot is 5%, calculate the probability of accepting the lot.

(ii) What is the Operating Characteristic (OC) Curve of this sampling plan? (06 *marks*)

Part II

5. (a) Describe the classical approach to probability. State **two** limitations of this approach to probability. (03 *marks*)

(b) A company has 40 female employees and 60 male employees. If two of them are selected at random what is the probability that

(i) both will be males?

(ii) both will be females?

(iii) there will be one female and one male?

(iv) Are these events collectively exhaustive and mutually exclusive? (05 *marks*)

(c) (i) Explain what you mean by conditional probability.
Why is the Bayes' theorem considered as a special case of conditional probability?

(ii) The probabilities of a man coming to work by train, bus, motor car or other means of transport are $\frac{3}{10}, \frac{2}{5}, \frac{1}{10}$ and $\frac{1}{5}$ respectively. The probabilities that he will be late if he comes by train, bus, motor car are $\frac{1}{4}, \frac{1}{3}$ and $\frac{1}{12}$ respectively but if he comes by other means he will not be late. If he is late for the work, what is the probability that he comes by train? (07 *marks*)

(d) Define the independence of two events A and B. Can mutually exclusive two events be independent?

The probability that the student A solving a problem is $\frac{3}{7}$ and the probability that the student B solving this problem is $\frac{7}{15}$.

(i) What is the probability that the problem will be solved, if both of them try it independent of each other?

(ii) What is the probability that **none** of them is able to solve the problem? (05 *marks*)

6. (a) Describe the random experiment relevant to the binomial distribution.

A multiple choice test consists of 10 questions and 4 answers to each question of which only one is correct. If a student who is completely unprepared for the test selects one answer randomly out of four for each question, find the probability of getting,

(i) exactly 3 correct answers.

(ii) at most 3 correct answers.

(iii) If the student gets 80% correct answers he will receive a distinction. What is the probability that the student will receive a distinction? (05 *marks*)

(b) Describe the random experiment relevant to the Poisson distribution.

Suppose that flaws in a certain fabric occur at random with an average of one per 20 square metres. What is the probability that a 2×5 metre piece of fabric will have

(i) no flaws?

(ii) at most one flaw?

(iii) at least two flaws? (05 *marks*)

[see page four

(c) The weights of bars of chocolates made in a factory are normally distributed with mean μ and variance σ^2. According to the past experience, 5% of bars weighted less than 90 grammes and $2\frac{1}{2}$% of bars weighted more than 100 grammes. Find the values of μ and σ^2 of the distribution and the percentage of bars produced which would be expected to weight less than 85 grammes. (05 *marks*)

(d) (i) State the conditions under which binomial distribution can be approximated by a normal distribution.

(ii) In a certain population 6% of the people have a certain disease. If 300 persons are selected at random from this population, what is the probability that more than 25 persons have the disease? (05 *marks*)

7. (a) In a small population with N=5 the y_i values are 2, 3, 6, 8, 11

(i) Calculate the population mean and population variance.

(ii) Using all the possible simple random samples of size 2 from this population, construct the sampling distribution of the sample mean $\bar{y}$ and show that $\bar{y}$ is an unbiassed estimator for population mean.

(iii) Find the variance of the sample mean $\bar{y}$ from the sampling distribution and verify that variance of $\bar{y}$ can be found using a formula with only a single sample since the population variance is known. (10 *marks*)

(b) Describe the method of systematic sampling. Explain **two** advantages and **two** disadvantages of systematic sampling. How do you compare the systematic sampling with cluster sampling? (05 *marks*)

(c) It is expected to estimate the percentage of rental houses in an area with 4000 houses within the range P$\pm$5 except one in twenty sample. Find the size of the sample that should be taken from the population to achieve this. You can assume the percentage of rental houses as 50%. (05 *marks*)

8. (a) A company is considering two different advertisements for the promotion of a new product. Advertisement A is used in one area and advertisement B is used in another area. In a random sample of 60 customers who saw the advertisement A, 36 bought the product. In a random sample of 80 customers who saw the advertisement B, 34 bought the product.

Find 95% confidence interval for the difference between true population proportions and state which advertisement is more effective. (06 *marks*)

(b) The mean and variance of the IQ values of a random sample of 40 boys were 98 and 160 respectively.

(i) Find a 99% confidence interval for the true mean of IQ values of the population.

(ii) Test the hypothesis that the true mean of the IQ values of the population is 100 at 5% significance level. (06 *marks*)

(c) A certain drug is claimed to be effective in curing colds. In an experiment on 200 people with cold, half of them were given the drug and other half of them were given sugar pills. The patients' reaction to the treatment are recorded in the following table.

	Helped	**Harmed**	**No effect**
Drug	60	15	25
Sugar Pills	50	10	40

Test the hypothesis that "drug is not better than sugar pills for curing colds" at 1% significance level. (08 *marks*)

* * *