.THE OPEN UNIVERSITY OF SRI LANKA
FACULTY OF HEALTH SCIENCES
DEPARTMENT OF BASIC SCIENCES
ACADEMIC YEAR 2019/2020 – SEMESTER 01

BACHELOR OF PHARMACY HONOURS
BACHELOR OF MEDICAL LABORATORY SCIENCES HONOURS
BSU4230 – BASIC STATISTICS – LEVEL 4
FINAL EXAMINATION
DURATION: TWO HOURS

DATE: 16th SEPTEMBER 2020   TIME: 9.30 am – 11.30 am

INDEX NO: ....

IMPORTANT INSTRUCTIONS / INFORMATION TO CANDIDATES

- This question paper consists of **07 pages** with **04 Essay Questions.**
- Necessary Formulae/ Z table are given in the pages 6-7.
- Write answers for **ALL** questions in the answer books provided.
- Write your **Index Number** in the space provided.
- Do NOT bring in or have in possession unauthorized materials, including mobile phones and other electronic devices, and do not violate any other examination rules.
- **Non-programmable calculators are allowed to use.**
- Do NOT take any page/part of this question paper out from the examination hall.

Essay Questions
(Total marks: 200)

1. In a study on obesity, a researcher selected a sample of 270 females and recorded the **height** (in *cm*), **weight** (in *kg*), **type of physical activity** over the past two years (recorded as: no physical activity, stretching, aerobics, other), **level of physical activity** (recorded as: inactive, somewhat active, fairly active, very active), **family history** of obesity (recorded as 1: yes; 0: no) and **age** (recorded as 1: less than 30 years, 2: 31 to 50 years, 3: above 50 years).

   i) Classify each of the six variables measured in this study as qualitative or quantitative.

   ii) Classify the quantitative variables as discrete or continuous.

   iii) Name the scale of measurement of each of the six variables as nominal, ordinal, interval or ratio.

   iv) State whether the study described here is an observational study or an experimental study. Give reasons for your answer.

   v) State whether the study described here is a retrospective study or a prospective study. Give reasons for your answer.

   (30 marks)

   vi) The following summary table presents the number of females in the sample falling into each cell in a 3 x 4 table, classified by age and level of physical activity.

Age	Level of physical activity			
	Inactive	Somewhat active	Fairly active	Very active
less than 30 years	24	36	55	45
31 to 50 years	23	38	15	04
above 50 years	14	11	04	01
Total	61	85	74	50

   a) Construct a relative frequency table to indicate how the sample is distributed across the three age categories.

   b) Suggest two suitable graphs that can illustrate the information in the relative frequency distribution constructed in part (a). **You need not construct the graphs.**

   c) **Construct** a suitable graph to highlight how the level of physical activity had varied with the age.

   d) Clearly describe five findings from the graph constructed in part (c).

   (20 marks)

2) i) In each of the studies given below, with appropriate reasoning for your answer, state whether you will expect a change in the sampling error/ a change in the non-sampling error or neither, due to the proposed change by the researcher (highlighted in bold case letters). If you decide that there is a change, state whether the error you identified would decrease or increase. If you decide that there is no change, give reasons.

a) In data collection for a survey, a comprehensive **training programme was given** for the data collectors prior to sending them for field work.

b) Having learnt from the pilot survey that the time taken to clot the blood in warfarin users in his population has a much larger variation than expected earlier, the researcher **decided to increase the sample size** from 150 to 350.

c) In a study to find out whether there is a gender difference in the mean times required to draw blood samples from elderly persons in the age group of 65 to 70 years, a researcher planned to take 250 persons from each group. However, due to inadequate number of males in the said age group arriving the laboratory for blood tests on the day of data collection, researcher decided to **include measurements collected on 100 males in the age group of 60 to 65 to have the required sample size and ignore the age in the analysis.**

(15 marks)

ii) A course administrator interested in estimating the average time that a student following the course BLU4335 spend on reading the course material, planned to collect data from 100 students who have registered for the course in the academic year 2019/2020. Of the total 450 students registered for the course in the academic year 2019/2020, 300 are employed and 150 are unemployed. The administrator suspects that employed and unemployed students differ with respect to the time spent on reading.

a) Define the terms population, sample and sampling unit in relation to this study.

(15 marks)

b) Suppose the course administrator seeks your advice to design the study. Clearly describe how you advise, outlining the main steps involved. Use the random number table given below to illustrate the sampling, selecting random numbers row-wise starting from the number 27569 marked in the table.

(20 marks)

Table of Random Numbers

```
05542 29705 83775 21564 81639 27973 62413 85652
19428 88048 08747 20092 12615 35046 67753 69630
97402 27569 90184 02338 39318 94936 34641 55525
51472 65358 23701 75230 47200 78176 85248 50589
98729 76703 16224 97661 79907 06611 26501 93389
61348 88857 53204 86721 59613 67494 17292 94457
93129 40386 51731 44254 66665 72835 01270 42523
59208 95266 33978 70958 60017 39723 00606 17956
96997 55340 80312 78839 09815 16887 22228 06206
```

3) a) A researcher has observed that 4 out of the 20 females sampled from residents in a female elderly home suffer from mild dementia. Suppose there are 200 females residing in this elderly home.

   i) Estimate the proportion of females in the elderly home suffering from mild dementia.

   ii) Estimate the standard error of the estimate calculated in part (i).

   iii) Estimate the total number of females in the elderly home suffering from mild dementia.

(20 marks)

   b) Of the 20 females in the sample described in part (a), 15 were above 70 years of age. Furthermore, out of the 4 females in the sample identified with mild dementia, 3 were from the age group of above 70 years. The researcher is interested in studying possible association between age and mild dementia, grouping age categorized as "above 70 years" and "at most 70 years" and mild dementia categorized as "identified" and "not identified".

   i) Construct a 2x2 contingency table that can be used to study the association between age and being identified with mild dementia.

   ii) Calculate the odds that a person is identified as mild dementia in the age group of above 70 years.

   iii) Calculate the odds ratio considering the two groups as at most 70 years and above. 70 years and interpret it in relation to this study.

(30 marks)

4. a) The following table was constructed using maternal ages of 100 mothers at the birth of their first child, reported in a rural region of Sri Lanka.

Age of the mother	Number of mothers
15 – 19	14
20 – 24	44
25 – 29	20
30 – 34	12
35 – 39	8
40 – 44	2

i) Calculate the sample mean and clearly explain the parameter that it estimates in relation to this study.

ii) The middle ten observations, after arranging the maternal ages of the 100 mothers in ascending order are 20, 20, 20, 20, 20, 20, 21, 21, 22 and 23. Calculate the median of the ten observations and clearly describe what one can conclude from this measure, in relation to this study.

iii) Based on the observations given in part (ii), a student concluded that the mode of the maternal ages of the 100 mothers in the sample is at least 6. Do you agree with this conclusion? Give reasons for your answer.

(20 marks)

b) Suppose birth weights of Sri Lankan baby girls is normally distributed with a mean of 3200 grams and a standard deviation of 500 grams.

i) Estimate the percentage of baby girls in the said population with birth weight exceeding 3300grams.

ii) Calculate the 10th percentile of the distribution of birth weights of baby girls in Sri Lanka.

iii) Clearly explain what you conclude about the birth weights of baby girls in Sri Lanka, based on the value calculated in part (ii).

(30 marks)

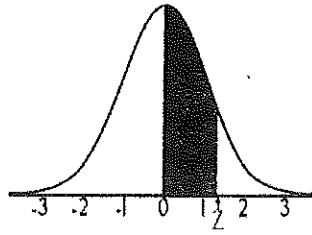***** Copyrights reserved *********

The following equations are given in the usual/ standard notation. You need to select the appropriate formulae from the given set of equations.

$$V = \frac{\sum (x - \bar{x})^2}{n - 1}$$

$$CV = \frac{S}{\bar{x}} \times 100\%$$

$$SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$$SE_p = \sqrt{\frac{pq}{n}}$$

STANDARD NORMAL TABLE (Z)

Entries in the table give the area under the curve between the mean and z standard deviations above the mean. For example, for z = 1.25 the area under the curve between the mean (0) and z is 0.3944.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0190	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2969	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3513	0.3554	0.3577	0.3529	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998