00922

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

BACHELOR OF SCIENCE HONOURS IN NURSING
BSU5335 – HEALTH STATISTICS - LEVEL 5
FINAL EXAMINATION
DURATION: 3 HOURS

DATE: 09th SEPTEMBER 2020 TIME: 09.30 AM – 12.30 PM

INDEX NO: .....

IMPORTANT INSTRUCTIONS/ INFORMATIONS TO CANDIDATES

- This question paper consists of **9 pages** with **06 Essay Questions.**
- **Each question carries 100 marks** and that corresponds to a **total of 600.**
- Write your **Index Number** in the space provided.
- Answer **ALL** the questions.
- **Questions should be answered in the booklet provided.**
- Necessary formulae/tables are given in the pages 7-9.
- Do NOT bring in on person or have in possession unauthorized materials, including mobile phones and other electronic devices, or violate any other examination rules.
- **Non-programmable calculators are allowed to use.**
- Do NOT remove any page/part of this question paper from the examination hall.

Essay Questions
(600 Marks)

1. a)

A random sample is selected from a population, and gender, age (to the nearest year as of 01st of January 2020), and the level of smoking (as never, occasionally, frequently) of the participants were recorded.

i) Name the scale of the measurement (nominal, ordinal, interval or ratio) of gender, age and level of smoking

ii) The age distribution of the participants is as follows.

Age group	Number of persons
20 – 30	10
31 – 41	32
42 – 52	30
53 – 63	8

Find the relative cumulative frequency corresponding to the third-class interval and describe what it measures in relation to this study.

(40 marks)

b) In a population, 38% are adult males, 42% are adult females and 20% are children. Prevalence of cancer among adult males, adult females and children in this population are 12%, 10% and 2% respectively. A person is randomly selected from this population.

Find:

i) the probability that the person has a cancer,

ii) the probability that the person is a child and has a cancer,

iii) the probability that the person is not an adult male and does not have a cancer,

iv) the probability that the person is an adult female, given that the person has a cancer.

(60 marks)

2.

a) In a large population, 20% of persons are infected with a certain disease. A random sample of 10 persons was examined and let X denote the total number of infected persons in the sample.

i) State the distribution of X.
ii) Find the probability that 8 persons were infected out of the 10 randomly selected persons.

(20 marks)

b) In a study conducted to investigate the death rate of a certain hospital in a rural area, a medical research group found that, on the average 49 deaths occur per week.

i) State a distribution that is suitable to model the number of deaths per week.
ii) Give reasons for the choice of the distribution stated in part (i).
iii) State the variance of the distribution given in part (i) and what it measures in relation to this study.
iv) Find the probability of
    a) 53 deaths in a given week.
    b) 5 deaths per day
    c) at least 2 deaths per day

(80 marks)

3.

a) In each of the following studies, state whether the researcher has selected probability sampling or non-probability sampling. Give reasons for your answer and write down one advantage and one disadvantage of the sampling method chosen by the researcher for each of the study.

i) To estimate the number of persons in an area infected with a certain disease, a researcher clinically tested all the persons who had symptoms of the disease.

ii) To estimate the total number of persons in Sri Lanka infected with the novel corona virus by PCR testing, a researcher randomly sampled persons from each and every district, determining sample sizes based on the population sizes of the districts.

iii) To find out the attitude of patients coming to a clinic about the facilities provided, a researcher collected data on 20th March 2020 by interviewing the 4th patient who arrived the clinic and every 5th patient thereafter, until he found the required sample size.

(50 marks)

b) To estimate the number of animals in a farm infected with a disease, a researcher plans to sample 50 animals from the 2500 animals in the farm. Suppose the disease gets spread through contacts and animals move freely within the farm so that there is frequent close contact among them.

i) Suppose the researcher wants to sample the animals on the same day and seeks your advice on how to do the sampling. Giving reasons for your choice, state the sampling method you would suggest to use.

ii) Now suppose the disease gets spread through sharing food and animals in the farm are fed in groups of 10. Will you still do the sampling as proposed in part (i) above? If not, clearly describe the changes you make.

iii) In a sample of 50 animals, suppose the researcher found 22 animals that are infected with the disease. Calculate the standard error of the proportion of infected animals. (answer should be correct up to 3 decimal places)

iv) Construct a 95% confidence interval for the population proportion. ($z = 1.96$)

(50 marks)

4.

a) State one advantage and one disadvantage of a cohort study over a case-control study.

(10 marks)

b) A random sample of 50 persons were selected from a population of 800 persons and based on a diagnostic test, 2 of the persons in the sample were found to be infected with a disease. If the diagnostic test has less sensitivity, would you expect the proportion of infected persons in the population to be less than 0.04, equal to 0.04 or more than 0.04. Give reasons for your answer.

(15 marks)

c) A researcher selected a random sample of 500 persons from a population and asked them to report number of visits they made to a hospital during the previous year to consult a doctor regarding cough problems (recorded as never or at least one visit) and exposure to dust at work place (recorded as regularly exposed, not regularly exposed).

i) State whether the researcher has conducted a case-control, cross-sectional or a cohort study.

(10 marks)

ii) Suppose in the sample of 500 persons, 100 reported as having regular exposure to dust and the rest reported as not having regular exposure to dust, at the workplace. Of these 100 persons who had regular exposure to dust, 20 had visited a doctor and the rest had not made any doctor visits, for cough problems. Of the 400 persons who had no regular exposure, 40 had visited a doctor for cough problems and the rest had not. Summarize the data in a 2×2 contingency table.

(10 marks)

iii) Calculate the odds of having to visit a doctor for a cough problem in the group who had regular exposure to dust at the workplace.

(15 marks)

iv) Calculate the odds ratio for the data summarized in part (ii) and interpret it.

(40 marks)

5.

a) A medical student conducted a study involving 10 patients to investigate the association between cholesterol level in blood (in milligrams per deciliter, mg/dL) and their maximum heart rate (beats per minute). In this study, heart rate was considered as the dependent variable (Y), and the cholesterol level was considered as the independent variable(X). Data are given below.

Cholesterol Level (X)	Maximum Heart Rate (Y)
239	99
211	88
265	109
149	63
204	85
230	95
164	69
303	125
269	111
165	70

$$\sum(x_i - \bar{x})^2 = 23473.3 \ , \ \sum(y_i - \bar{y})^2 = 3730$$

i. Calculate $\bar{x}$ (mean value of x variable) and $\bar{y}$ (mean value of y variable).

ii. Calculate the Pearson correlation coefficient using the given information.

iii. Clearly describe what you can conclude about the relationship between the variables based on the value calculated in part (ii)

(60 marks)

b) Suppose the medical student further wants to predict the maximum heart rate based on the cholesterol level and wants to fit a least squares linear regression line for the above data. (**Hint:** Least squares regression line is defined as $Y = a + bX$)

i) Calculate the slope of the fitted line.

ii) Calculate the intercept of the fitted line.

iii) Write down the fitted regression equation.

iv) Using the fitted regression equation, estimate the maximum heart rate of a patient, whose cholesterol level is 200 mg/dL.

(40 Marks)

6.

a) State one advantage and one disadvantage of using a parametric test over a non-parametric test to examine the validity of a hypothesis based on a statistical test.

(10 marks)

b) In a study on free thyroxin levels (T4) of female children of age 11 to 16 years, residing in a coastal area, the sample mean and the standard deviation of the thyroxin levels measured on 100 children were 12.8 pmol/l and 1.4 pmol/l respectively.

The researcher wants to know whether the mean free thyroxin level (T4) of the population is significantly different from 13.0 pmol/l or not.

i) Clearly describing your notation, state the null and the alternative hypotheses that you would examine to address the researcher's objective.

(10 marks)

ii) Clearly stating the assumptions, write down a suitable test statistic that can be used to test the null hypotheses stated in part (i), based on the given information.

(15 marks)

iii) Test the hypothesis using a 5% significance level and clearly state the conclusions you make regarding the objectives of the researcher.

(40 marks)

iv) State whether you would use the same statistical test that has been used in part (iii) to test the null hypotheses stated in part (i), in each of the following cases. If you use the same statistical test, clearly describe additional assumptions (if any) that you would make. If you use a different statistical test, clearly state the statistical test you would use; in each case, if any additional data other than the summary statistics provided are needed for the suggested test describe clearly.

a. The sample size is 24 and not 100.

b. A histogram of the data indicated that the distribution is skewed to the right.

(25 marks)

Necessary Formulae

The following equations are given in the usual/ standard notation.

Standard Error

$$SE(\bar{x}) = \frac{SD}{\sqrt{n}}$$

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Pearson Correlation Coefficient

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{[\sum(x-\bar{x})^2 \sum(y-\bar{y})^2]}}$$

Confidence intervals

$$\bar{x} \pm z\frac{SD}{\sqrt{n}}$$

$$\hat{p} \pm z\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Regression Parameters

$$b = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}$$

$$a = \bar{y} - b\bar{x}$$

Test statistic

$$z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$$

$$t = \frac{\bar{x}-\mu_0}{SD/\sqrt{n}}$$

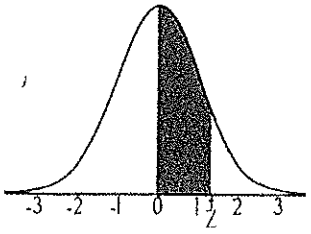
Distributions

$$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x} \;;\; x = 0,1,\ldots,n$$

$$\binom{n}{x} = \frac{n!}{(n-x)!\,x!}$$

$$P(X=x) = \frac{\lambda^x e^{-\lambda}}{x!} \;;\; x = 0,1,2,\ldots$$

P.T.O



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

t distribution table

df/p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
1	0.324920	1.000000	3.077684	6.313752	12.70620	31.82052	63.65674	636.6192
2	0.288675	0.816497	1.885618	2.919986	4.30265	6.96456	9.92484	31.5991
3	0.276671	0.764892	1.637744	2.353363	3.18245	4.54070	5.84091	12.9240
4	0.270722	0.740697	1.533206	2.131847	2.77645	3.74695	4.60409	8.6103
5	0.267181	0.726687	1.475884	2.015048	2.57058	3.36493	4.03214	6.8688
6	0.264835	0.717558	1.439756	1.943180	2.44691	3.14267	3.70743	5.9588
7	0.263167	0.711142	1.414924	1.894579	2.36462	2.99795	3.49948	5.4079
8	0.261921	0.706387	1.396815	1.859548	2.30600	2.89646	3.35539	5.0413
9	0.260955	0.702722	1.383029	1.833113	2.26216	2.82144	3.24984	4.7809
10	0.260185	0.699812	1.372184	1.812461	2.22814	2.76377	3.16927	4.5869
11	0.259556	0.697445	1.363430	1.795885	2.20099	2.71808	3.10581	4.4370
12	0.259033	0.695483	1.356217	1.782288	2.17881	2.68100	3.05454	43178
13	0.258591	0.693829	1.350171	1.770933	2.16037	2.65031	3.01228	4.2208
14	0.258213	0.692417	1.345030	1.761310	2.14479	2.62449	2.97684	4.1405
15	0.257885	0.691197	1.340606	1.753050	2.13145	2.60248	2.94671	4.0728
16	0.257599	0.690132	1.336757	1.745884	2.11991	2.58349	2.92078	4.0150
17	0.257347	0.689195	1.333379	1.739607	2.10982	2.56693	2.89823	3.9651
18	0.257123	0.688364	1.330391	1.734064	2.10092	2.55238	2.87844	3.9216
19	0.256923	0.687621	1.327728	1.729133	2.09302	2.53948	2.86093	3.8834
20	0.256743	0.686954	1.325341	1.724718	2.08596	2.52798	2.84534	3.8495
21	0.256580	0.686352	1.323188	1.720743	2.07961	2.51765	2.83136	3.8193
22	0.256432	0.685805	1.321237	1.717144	2.07387	2.50832	2.81876	3.7921
23	0.256297	0.685306	1.319460	1.713872	2.06866	2.49987	2.80734	3.7676
24	0.256173	0.684850	1.317836	1.710882	2.06390	2.49216	2.79694	3.7454
25	0.256060	0.684430	1.316345	1.708141	2.05954	2.48511	2.78744	3.7251
26	0.255955	0.684043	1.314972	1.705618	2.05553	2.47863	2.77871	3.7066
27	0.255858	0.683685	1.313703	1.703288	2.05183	2.47266	2.77068	3.6896
28	0.255768	0.683353	1.312527	1.701131	2.04841	2.46714	2.76326	3.6739
29	0.255684	0.683044	1.311434	1.699127	2.04523	2.46202	2.75639	3.6594
30	0.255605	0.682756	1.310415	1.697261	2.04227	2.45726	2.75000	3.6460
z	0.253347	0.674490	1.281552	1.644854	1.95996	2.32635	2.57583	3.2905
CI	———	———	80%	90%	95%	98%	99%	99.9%