THE OPEN UNIVERSITY OF SRI LANKA
FACULTY OF HEALTH SCIENCES
DEPARTMENT OF BASIC SCIENCES
ACADEMIC YEAR 2018/2019 – SEMESTER II

BACHELOR OF PHARMACY HONOURS – LEVEL 5
BACHELOR OF MEDICAL LABORATORY SCIENCES HONOURS – LEVEL 5
BSU5230 – APPLIED STATISTICS
FINAL EXAMINATION **DURATION: TWO HOUR**

DATE: 25th November 2019 **TIME: 9.30 am – 11.30 am**

INDEX NO: .....

IMPORTANT INSTRUCTIONS/ INFORMATIONS TO CANDIDATES

- This question paper consists of 09 pages with 4 Essay Questions.
- Write your **Index Number** in the space provided.
- Answer **ALL questions in the booklet provided.** (All answers must be in **INK:** Answers in pencil will **NOT** be marked))
- Necessary Formulae/ tables are given in page 06-09.
- Mobile phones and any other electronic equipment are **NOT** allowed.
- **Non programmable Calculators are allowed.**

Essay Questions (100% Marks)

01.

a) A health research firm wishes to estimate the percentage of working women having high blood pressure.
A sample of 625 working women in a city were selected and tested to determine whether they have high blood pressure. Out of the 625 women tested, 500 turned out to have high blood pressure.

i. Explain the difference between point estimates and interval estimates.

ii. What is the point estimate of the percentage of working women with high blood pressure?

iii. Calculate the standard error.

iv. Calculate 95% and 99% confidence intervals for the population proportion of working women having high blood Pressure.

v. Interpret the 99% confidence interval that you have calculated in part iv above.

(25 marks)

b) A study was designed to test the claim "the population proportion of male students who consume artificial protein supplements is 0.6".
A random sample shows that 77 out of 100 male students surveyed, consume artificial protein supplements. The study group wishes to test the above claim at 0.05 significance level.

i. Clearly state the null hypothesis and the alternative hypothesis.

ii. Calculate the test statistic value.

iii. What is the distribution and the corresponding table value that can be used to test the null hypothesis?

iv. Can we reject the null hypothesis at 0.05 significance level? Give reason/s.

v. What is your conclusion about the population proportion of males who consume artificial protein supplements?

(25 marks)

02. A group of lab scientists aim to investigate whether there is an association in cure rates between males and females for a specific injection treatment. The following table provides the observed frequencies of cured and non-cured patients for both genders.

	Males	Females	Row total
Cured	12	48	60
Not-cured	30	10	40
Column total	42	58	100

i. Clearly state the null hypothesis and the alternative hypothesis.

ii. Calculate the expected frequencies for these observed frequencies.

iii. Calculate the Chi square test statistic

iv. What is the degree of freedom value for this test?

v. Find the corresponding table value at 0.05 significance level.

vi. What is your conclusion about the cure rates between males and females at 0.05 significance level? Give reason/s. **(50 marks)**

03.

a) In a Pharmaceutical industry, a researcher wishes to test whether mean weight of an active ingredient of a particular tablet is different from the standard value of 13.00 milligrams. The masses, in milligrams, of 10 randomly selected tablets are given below. Standard deviation of the sample is 1.17 (Assume that this sample comes from a normal population)

Number	1	2	3	4	5	6	7	8	9	10
Weigh (milligrams)	12.50	13.80	14.60	12.30	14.10	14.00	11.10	13.30	12.50	14.80

i. Clearly state the null hypothesis and alternative hypothesis.

ii. What is the suitable parametric test that the researcher may use? Give reason/s.

iii. What is the average weight of the sample?

iv. Calculate the test statistic.

v. What is the degrees of freedom value and the corresponding table value at 0.05 significance level for the above test?

vi. What is your conclusion about the mean weight of active ingredient of the tablets? Give reason/s. **(25 marks)**

b) A new prenatal care program is introduced for women who are living in a rural area. It involves in-home visits during the course of pregnancy in addition to the usual scheduled visits. Out of a sample of 15 pregnant women, 7 are randomly assigned for the new prenatal care program and others participated in the usual care program. After delivering their babies, APGAR scores were e measured to see whether women, who participate in the new program, deliver healthier babies than women receiving only usual care. (APGAR score is a measure of the physical condition of a newborn infant which stands for "Appearance, Pulse, Grimace, Activity, and Respiration). APGAR scores range from 0 to 10 with scores of 7 or higher considered as normal (healthy), 4-6 is low and 0-3 is critically low. The data are shown below.

Usual Care	8	7	6	2	5	8	7	3
New Program	9	9	7	8	10	9	6	

i. What is the difference between parametric test and non-parametric test?

ii. What is the suitable non-parametric test to see whether there is any significant difference between APGAR scores between babies in the usual prenatal care and the new prenatal care program?

iii. Clearly state the null hypothesis and alternative hypothesis.

iv. Perform the test that you mention in part ii and interpret the result. (Hint: Table values at 5% significance level are $lower\ limit = 41$ and $upper\ limit = 71$)

(25 marks

04.

a)

i. Explain the following terms

A) Pearson's correlation coefficient
B) Regression

ii. Explain the difference between simple linear regression and multiple linear regression.

iii. Sketch the scatter plots for the following correlation coefficient values and interpret them.

A) 0.50
B) 0.85
C) -1.0

iv. Identify the symbols of the linear regression line $\hat{Y} = bX + a$. **(25 marks)**

b) A researcher wishes to compare three drugs used to treat some disease. Responses (time to cure in days) are measured such that a smaller value indicates a more favorable response. A total of 18 patients are randomly assigned to one of the three drug groups (Drug A, Drug B and Drug C). The data are provided below.

Drug A	Drug B	Drug C
7.30	7.10	5.80
8.20	10.60	6.50
10.10	11.20	8.80
6.00	9.00	4.90
9.50	8.50	7.90
	10.90	8.50
		5.20

i. State the null hypothesis and alternative hypothesis to be tested.

ii. Calculate the average cure times in each treatment separately.

iii. Calculate overall average cure time for all treatments together.

iv. Calculate the F-test statistic using summarized data given in the following table.

	Drug A	Drug B	Drug C
Sample variance	2.73	2.61	2.56

v. If the critical F value at 0.05 of significance is 3.63, interpret the results. **(25 marks)**

Necessary Formulae

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

Test Statistics

$$T = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

$$T = \frac{p - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}}$$

$$Z = \frac{x_1 - x_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$Z = \frac{p_1 - p_2}{\sqrt{p(1-p)[\frac{1}{n_1} + \frac{1}{n_2}]}}$$

$$T = \frac{x_1 - x_2}{SD_D/\sqrt{n}}, \quad SD_D = \sqrt{\frac{\sum D^2}{n-1} - \bar{x}_d^2}$$

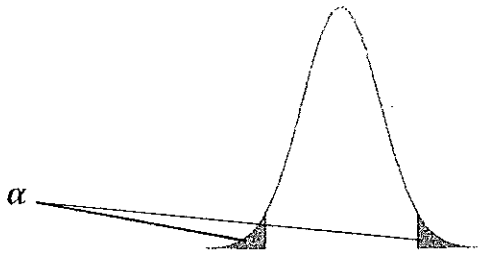
$$\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e}$$

$$(Expected\ frequency = \frac{row\ total \times column\ total}{grand\ total})$$

$$F = \frac{MST}{MSE}$$

$$MST = \frac{n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \ \ldots\ldots\ldots + n_k(\bar{x}_k - \bar{x})^2}{k - 1}$$

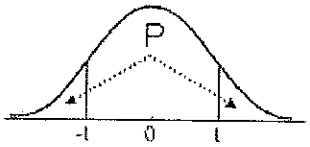
$$MSE = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2 + \ \ldots\ldots\ldots + (n_k - 1)S_k^2}{n - k}$$



Z table values

Significance level (α)	Z value
0.01	2.58
0.05	1.96
0.10	1.64

t table



DF	0.20	0.10	0.05	0.02	0.01	0.005	0.002	0.001
1	3.078	6.314	12.706	31.820	63.657	127.321	318.309	636.619
2	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.599
3	1.638	2.353	3.182	4.541	5.841	7.453	10.215	12.924
4	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	1.397	1.860	2.306	2.897	3.355	3.833	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	1.345	1.761	2.145	2.625	2.977	3.326	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	1.337	1.746	2.120	2.584	2.921	3.252	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792

Chi Squared table

df	Significance levels		
	0.05	0.01	0.001
1	3.84	6.64	10.83
2	5.99	9.21	13.82
3	7.82	11.35	16.27
4	9.49	13.28	18.47
5	11.07	15.09	20.52
6	12.59	16.81	22.46
7	14.07	18.48	24.32
8	15.51	20.09	26.13
9	16.92	21.67	27.88
10	18.31	23.21	29.59
11	19.68	24.73	31.26
12	21.03	26.22	32.91
13	22.36	27.69	34.53
14	23.69	29.14	36.12
15	25.00	30.58	37.70
16	26.30	32.00	39.25
17	27.59	33.41	40.79
18	28.87	34.81	42.31
19	30.14	36.19	43.82
20	31.41	37.57	45.32