THE OPEN UNIVERSITY OF SRI LANKA
FACULTY OF HEALTH SCIENCES
DEPARTMENT OF BASIC SCIENCES
ACADEMIC YEAR 2023/2024 – SEMESTER 02

BACHELOR OF PHARMACY HONOURS
BACHELOR OF MEDICAL LABORATORY SCIENCES HONOUR
BSU5230 – APPLIED STATISTICS – LEVEL 05
FINAL EXAMINATION
DURATION: 2 HOURS

DATE: 24th October 2024 TIME: 1.30PM – 3.30PM

REGISTRATION NO: .....

IMPORTANT INSTRUCTIONS / INFORMATION TO CANDIDATES

- This question paper consists of **08 pages** with **04 Essay Questions**.
- Answer **ALL** questions.
- Write your **Index Number** in the space provided.
- **All the answers should be answered in the booklet provided.**
- Necessary statistical tables and equations are given in **pages 6-8.**
- **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.
- **DO NOT** removes any page/part of this question paper from the examination hall.
- **Non-programmable calculators are allowed.**

51

Essay Questions

(200 marks)

1. A study was conducted to explore the relationship between smoking status and lung cancer incidence in a population of 200 individuals. The participants were categorized as either "smokers" or "non-smokers," and their subsequent diagnosis of lung cancer was recorded as either "positive" or "negative." The data collected is summarized in the contingency table given below:

Smoking Status	Lung Cancer Diagnosis	
	Positive	Negative
Smokers	40	60
Non-smokers	20	80

i. Explain the concept of 'chi-squared test' in statistical analysis. **(06 marks)**

ii. State the null hypothesis and alternative hypothesis relevant to the above study. **(06 marks)**

iii. Determine the values for the letters A-H given in the following table. **(24 marks)**

Observed Value (f_0)	Expected Value (f_e)	$\frac{(f_0 - f_e)^2}{f_e}$
40	A	E
60	B	F
20	C	G
80	D	H

iv. Calculate the chi-squared test statistic. **(04 marks)**

v. What is the degrees of freedom value and the corresponding critical value (table value) at 0.05 significance level for the above test. **(04 marks)**

vi. Compare the test statistics with critical value and interpret the results. **(06 marks)**

2.

a)

i. Briefly describe what is hypothesis testing? **(03 marks)**

ii. Write down main 4 steps in hypothesis testing. **(04 marks)**

iii. Briefly explain the difference between descriptive statistics and inferential statistics. **(03 marks)**

b) A researcher aims to assess the impact of a new exercise program on weight reduction in a group of 10 participants. Before starting the exercise program, weight of the participants were recorded. After six weeks of following the exercise program, their weight was measured again. The data collected is presented below (Assume data are normally distributed).

Participant	Initial Weight (kg) (IWB)	Weight after 6 weeks (kg) (IWA)
1	85	80
2	72	68
3	91	86
4	78	75
5	67	65
6	95	90
7	82	78
8	70	67
9	88	84
10	75	72

i. What is the suitable parametric test that the researcher could use? **(02 marks)**

ii. Clearly State the Null hypothesis and Alternative hypothesis. **(04 marks)**

iii. Complete the following table (Copy the table to your answer booklet). **(20 marks)**

Participant	IWB	IWA	Difference (D)	D^2
1	85	80		
2	72	68		
3	91	86		
4	78	75		
5	67	65		
6	95	90		
7	82	78		
8	70	67		
9	88	84		
10	75	72		

iv. Calculate the test statistic. **(08 marks)**

v. Obtain the critical value (table value) at 5% significance level. **(03 marks)**

vi. What is your conclusion? **(03 marks)**

3.

a)

i. List 2 differences of parametric and non-parametric tests? **(04 marks)**

ii. Explain nature of data used in the following non-parametric tests. **(06 marks)**

1) Wilcoxon Signed-Rank test
2) Wilcoxon Rank-Sum test
3) Spearman's Rank Correlation test

b) A researcher aims to explore the relationship between physical exercise and reduction in cholesterol levels among 5 participants. The data collected includes the number of hours each participant exercised per week and their corresponding reduction in cholesterol levels after six weeks. The data is presented in the following table

Participant	Hours of Exercise per Week (X)	Cholesterol Reduction (mg/dL) (Y)
1	5	15
2	7	22
3	4	10
4	6	18
5	8	25

i. Draw a scatterplot and describe the nature of the relationship between the two variables based on that. **(10 marks)**

ii. Complete the following table (Copy the table to your answer booklet) **(15 marks)**

X	Y	$(X-\bar{X})^2$	$(Y-\bar{Y})^2$	$(X-\bar{X})(Y-\bar{Y})$
5	15			
7	22			
4	10			
6	18			
8	25			

iii. Calculate the Pearson Correlation Coefficient. **(10 marks)**

iv. Using the calculated Pearson correlation coefficient, discuss whether there is a linear association between the hours of exercising and cholesterol reduction. **(05 marks)**

4.

a)

i. Describe what is Analysis of Variance (ANOVA)? **(06 marks)**

ii. Write down main 3 assumptions of ANOVA. **(06 marks)**

iii. Clearly State the common Null hypothesis and Alternative hypothesis for ANOVA testing. **(08 marks)**

c) A researcher is investigating the impact of study hours on students' exam scores. The following data was collected from 10 students, the number of hours they studied and their exam scores. The data is provided below.

Student	Hours Studied	Exam Score
1	2	50
2	4	55
3	6	60
4	8	65
5	10	70
6	12	75
7	14	80
8	16	85
9	18	90
10	20	95

i. Identify the dependent and independent variables. **(06 marks)**

ii. Estimate the Regression Equation. **(20 marks)**

iii. Using the regression equation, predict the exam score of a student who has studied 15 hours. **(04 marks)**

List of Equations

$$t = \frac{(\Sigma D)/N}{\sqrt{\frac{\Sigma D^2 - \left(\frac{(\Sigma D)^2}{N}\right)}{(N-1)(N)}}}$$

$$r_{xy} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma(x_i - \bar{x})^2 \Sigma(y_i - \bar{y})^2}}$$

r_{xy} = correlation coefficient between x and y

x_i = the values of x within a sample

y_i = the values of y within a sample

$\bar{x}$ = the average of the values of x within a sample

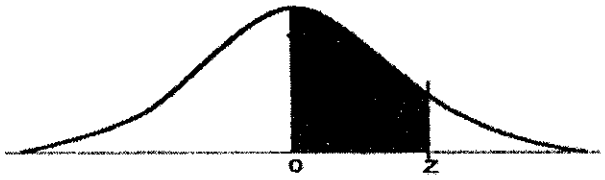
$\bar{y}$ = the average of the values of y within a sample

$$\beta_1 = \frac{\sum_{i=1}^{m}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{m}(x_i - \bar{x})^2}$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

Chi-square Distribution Table

d.f.	.995	.99	.975	.95	.9	.1	.05	.025	.01
1	0.00	0.00	0.00	0.00	0.02	2.71	3.84	5.02	6.63
2	0.01	0.02	0.05	0.10	0.21	4.61	5.99	7.38	9.21
3	0.07	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34
4	0.21	0.30	0.48	0.71	1.06	7.78	9.49	11.14	13.28
5	0.41	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09
6	0.68	0.87	1.24	1.64	2.20	10.64	12.59	14.45	16.81
7	0.99	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.72
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22



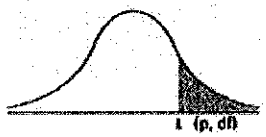
This table presents the area between the mean and the Z score . When Z=1.96, the shaded area is 0.4750.

Areas Under the Standard Normal Curve

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	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.6	.4998	.4998	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.9	.5000									

Source: Adapted by permission from *Statistical Methods* by George W. Snedecor and William G. Cochran, sixth edit
© 1967 by The Iowa State University Press, Ames, Iowa, p. 548.

Numbers in each row of the table are values on a *t*-distribution with (*df*) degrees of freedom for selected right-tail (greater-than) probabilities (*p*).



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%