THE OPEN UNIVERSITY OF SRI LANKA
Faculty of Engineering Technology
Department of Mathematics and Philosophy of Engineering

031

Bachelor of Industrial Studies Honors

Final Examination (2022/2023)
MHZ4357: Applied Statistics

Index No: .....

Date: 19/02/2024 Time: 09:30-12:30 hours .

Instructions:

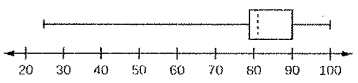
- **Part A is Compulsory**
  - Provide short answers in given space.
  - Do not need to show any workings.
- Answer five (05) questions only from Part B.
  - Provide answers in separate sheets (answer booklet) which will be given in the examination.
  - Show all your workings.
- Number of pages in the paper is Seven (07).
- All the symbols are in standard notation unless they are defined.
- All the relevant statistical tables are attached with this paper.
- Do not need to use graph sheets in this paper.
- **Attach the Part A to the Answer script of the Part B.**
- This is a closed book test and do not use red color pen.

Part A

Provide short answers in the given space. Do not need to show any workings or works. Please attach this part to your answer script.

1. Explain the difference between a Interval scale variable and a Ratio scale variable.
   Answer:.....
   ..... *(Marks 10)*

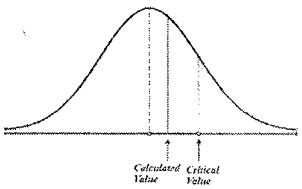
2. What can you say about the shape of the distribution of data, which is represented by the following Box and whisker Plot?
   Answer:.....



   ..... *(Marks 10)*

3. If K is the mean of a random variable (X) which follows the Poisson distribution, what is the standard deviation of the random variable X ?
   Answer:.....
   (Marks 10)

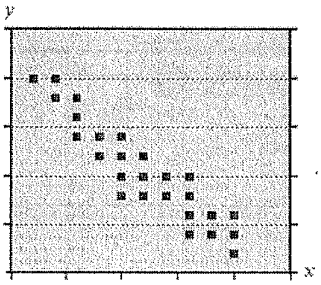
4. What is the decision that can be made according to the position of calculated and critical Z values as displayed in the graph when doing hypothesis testing?



Answer:.....
.....
.....

(*Marks 10*)

5. Consider the scatter plot given in following figure and identify the relationship between the two variables x and y.



Answer:.....
.....
.....

(*Marks 10*)

6. Let A and B be two independent events such that $P(A \cap B) = 1/18$. If it is given $P(A \cup B) = 4/9$ and B has a probability of occurring twice that of A. Then find the values of $P(A)$ and $P(B)$.

   Answer:..... (*Marks 10*)

7. Let the mean of the observations: x, 5, 6, 8, and 7 is given as 6. Then find the value of the x.

   Answer:.....

   (*Marks 10*)

8. What is the meaning of the level of significance in hypothesis testing?

   Answer:.....
   ..... (*Marks 10*)

9. Suppose that you want to calculate the 90% confidence interval for the mean value of a normally distributed population with a known variance $\sigma^2 = 4$ with a sample of 36 units. Find the standard of error of mean.

   Answer:..... (*Marks 10*)

10. A researcher wants to examine the impact of various level of fertilizer on yeild of the fruits (Kg per plant). Identify the response and predictor variables related to the above experiment.

    Answer:.....
    ..... (*Marks 10*)

Part B

1. A researcher interest to know the effect of type of fertilizers and experience of the farmer on the production of Tomatoes in Bandarawela. He has gathered the following data.

Type of the Fertilizer	Farming Experience(years)	Yield (Kg per Unit Area)
1	26	30.5
2	27	37.4
3	27	32.1
1	28	40.0
1	30	44.1
2	36	40.2
2	29	38.1
1	29	39.0
3	37	42.1
3	16	38.2
3	16	31.3
1	17	35.4
2	30	38.3
2	6	20.5
1	32	30.5

(a) Identify the type of data associated with the following variables in the given data set. *(Marks 15)*

i. Type of the Fertilizer

ii. Farming Experience (years)

iii. Tomato yield (Kg per area)

(b) State the scale of measurement of the following variables. *(Marks 10)*

i. Tomato yield (Kg per area)

ii. Type of the Fertilizer

(c) Draw an appropriate graph/chart to show the composition of sample among the type of Fertilizer. Interpret your findings in appropriate way. *(Marks 15)*

(d) Draw an appropriate graph/chart to evaluate the relationship among the farmer's experince on the Tomato yield (Kg per area) and interpret your findings in appropriate way.

(Marks 15)

(e) Compile a table of the frequency distribution and the relative frequency distribution to summerize the data in Tomatoe yield(Kg per area) where it is given that the number of classes is 4. *(Marks 20)*

(f) Draw the histogram for the above frequency distribution. *(Marks 15)*

(g) Comment about the shape of the distribution of the sample data. *(Marks 10)*

2. The following table gives the weight losses per 1kg in the process of food drying on a sample of 36 number of foods.

Weight loss(g)	10	10.5	11	11.5	12	12.5	13	14	14.5
Frequency	3	4	7	8	6	5	1	1	1

(a) Determine the central tendency measures (Mean, Median and Mode) of the above sample. *(Marks 20)*

(b) Find the varience and inter-quartile range for the sample data.
(Hint:$\sum_{i=1}^{36}(x-\bar{x})^2 = 36.25$) *(Marks 30)*

(c) Draw the box-whisker plot for the above data. *(Marks 20)*

(d) Are there any outliers in the given data set? Justify your answer. *(Marks 20)*

(e) What are the best suited measures for the central tendency? Explain your answer. *(Marks 10)*

3. In an experiment, it was found that the number of defective fruits identified in a quality checking machine, follows the probability mass function given in the following table. Let X be the random variable which represent the number of defective fruits identified in quality checking machine.

$X = x$	1	2	3	4	5	6	Otherwise
$P(X = x)$	0.1	0.2	0.1	0.3	0.1	k	0

(a) Identify the set of all possible outcomes that the random variable X can has. *(Marks 10)*

(b) Find the value of the k here. *(Marks 20)*

(c) Find the probability that it may identified more than 4 defective fruits with the qaulity checking machine. *(Marks 15)*

(d) Find the probability that it can be having more than two and less than or equal four defective fruits with the qaulity checking machine. *(Marks 15)*

(e) Find the expected number and the variance of defective fruits with the qaulity checking machine. *(Marks 40)*

4. (a) Discuss the characteristics of Geometric distribution and the characteristic of Binomial distribution.

(*Marks 20*)

(b) A fruit harvester machine is used to detect proper fruits for use. Previous studies confirm that the probability of a fruit being an unusable fruit is 0.03.

i. Find the probability that the first unusable fruit is obtained at the 5th inspected fruit. (*Marks 20*)

ii. What is the probability that the machine will inspect more than 3 fruits to detect the first defective item? (*Marks 20*)

iii. Let the same machine is used to identifyt the unusable fruits in a batch of 20 fruits. Find the probability that 10% of the batch of the fruits were unusable.

(*Marks 40*)

5. (a) Define the following terms.

i. Interval estimate (*Marks 10*)

ii. Magin of error of the statistic (*Marks 10*)

(b) An agricultural researcher plants 25 plots with a new variety of corn. The average yield for these plots is $\bar{x} = 150$ bushels per acre and the Standard deviation of the Yield in these plots is $s = 10$ bushels. Assume that the yield per acre for the new variety of corn follows a normal distribution with unknown mean μ.

i. Give an estimate for the average yield of the new variety of the corn plant.

(*Marks 10*)

ii. Find the standard error of the yield of the new variety of the corn plant.

(*Marks 15*)

iii. If the resercher wants to find a 90% confidence interval for the average yield of the new variety of the corn plant, then

A. Find the critical value associated with this caculation. Explain the reason for selecting the corresponing table. (*Marks 20*)

B. Find the margin of error for this calculation. (*Marks 10*)

C. Find the confidence Interval and interpret it in a meaningful way.

(*Marks 25*)

6. A electronic machine is programmed to packet the 5kg rice packets. However, customers had complained that the weight of the rice packets which are packets by the said machine are less than to 5kg. So, the programmer collected data of the weight of rice per packets from a sample of 36, 5kg packets of rice which are packets by the said machine, and identified that the sample mean weight of rice is 4.8kg and sample variance is 0.25kg per packet. Then the programmer use hypothesis testing to see the validity of the customers' claim.

   (a) State the corresponding null and alternative hypothesises that the programmer of the rice packeting machine test to see the validity of the customers' complain. (*Marks 20*)

   (b) Find the test statistics associated with this hypothesis testing. (*Marks 20*)

   (c) If the owner wants to test this hypothesis at 5% significant level, then what is the critical value/s associates with this test? (*Marks 15*)

   (d) Explain why did you choose the corresponding table when finding critical value/s in part (c). (*Marks 10*)

   (e) Test the hypothesis at 5% significant level and state your decision regarding the hypothesis. Explain your conclusion regarding the claim of the customer. (*Marks 35*)

7. A study was conducted to determine whether the number of acre harversted has an impact on the monthly compensation of farmers in Nuwara Elliya. The table below provides the number of acre harvested and monthly income of 10 farmers.

NUMBER OF ACRE HARVESTED	13	16	22	2	8	26	19	20	1	4
MONTHLY COMPENSATION IN THOUSAND RUPEES	260	330	360	160	260	370	330	360	160	190

   (a) Identify the independent variable and the dependent variable. (*Marks 10*)

   (b) Calculate the correlation coefficient for above data. Interpret your answer. (*Marks 25*)

   (c) Draw the scatter plot for dependent variable against the independent variable.(*Marks 10*)

   (d) Find the coefficient of the equation of the best fitted simple linear regression model of the above data. (*Marks 25*)

   (e) Interpret the found regression coefficient in part (d). (*Marks 10*)

(f) Use the estimated line of regression to find the monthly income of a farmer who has harvested in 10 acres. (*Marks 10*)

(g) State the assumptions that you have to make within this analysis. (*Marks 10*)

Important Formulars:

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

Binomial distribution

$$Pr(X = r) = \frac{n!}{(r!(n-r)!)} p^r q^{(n-r)}$$

Geometric distribution

$$Pr(X = r) = pq^{(r-1)}$$

Poison Distribution

$$Pr(X = r) = e^{(-\pi)} \frac{\mu^r}{r!}$$

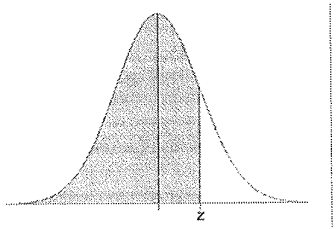
If random variable X, follows normal distribution ($X \sim N(\mu, \sigma^2)$) Then $Z = \frac{(X-\mu)}{\sigma}$ follows standard normal distribution where $Z \sim (1, 0)$.

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}$$

$$\beta_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$$

End.

Standard Normal Cumulative Probability Table



Cumulative probabilities for POSITIVE z-values are shown in the following table:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

t Distribution: Critical Values of *t*

		Significance level					
Degrees of freedom	*Two-tailed test:*	10%	5%	2%	1%	0.2%	0.1%
	One-tailed test:	5%	2.5%	1%	0.5%	0.1%	0.05%
1		6.314	12.706	31.821	63.657	318.309	636.619
2		2.920	4.303	6.965	9.925	22.327	31.599
3		2.353	3.182	4.541	5.841	10.215	12.924
4		2.132	2.776	3.747	4.604	7.173	8.610
5		2.015	2.571	3.365	4.032	5.893	6.869
6		1.943	2.447	3.143	3.707	5.208	5.959
7		1.894	2.365	2.998	3.499	4.785	5.408
8		1.860	2.306	2.896	3.355	4.501	5.041
9		1.833	2.262	2.821	3.250	4.297	4.781
10		1.812	2.228	2.764	3.169	4.144	4.587
11		1.796	2.201	2.718	3.106	4.025	4.437
12		1.782	2.179	2.681	3.055	3.930	4.318
13		1.771	2.160	2.650	3.012	3.852	4.221
14		1.761	2.145	2.624	2.977	3.787	4.140
15		1.753	2.131	2.602	2.947	3.733	4.073
16		1.746	2.120	2.583	2.921	3.686	4.015
17		1.740	2.110	2.567	2.898	3.646	3.965
18		1.734	2.101	2.552	2.878	3.610	3.922
19		1.729	2.093	2.539	2.861	3.579	3.883
20		1.725	2.086	2.528	2.845	3.552	3.850
21		1.721	2.080	2.518	2.831	3.527	3.819
22		1.717	2.074	2.508	2.819	3.505	3.792
23		1.714	2.069	2.500	2.807	3.485	3.768
24		1.711	2.064	2.492	2.797	3.467	3.745
25		1.708	2.060	2.485	2.787	3.450	3.725
26		1.706	2.056	2.479	2.779	3.435	3.707
27		1.703	2.052	2.473	2.771	3.421	3.690
28		1.701	2.048	2.467	2.763	3.408	3.674
29		1.699	2.045	2.462	2.756	3.396	3.659
30		1.697	2.042	2.457	2.750	3.385	3.646
32		1.694	2.037	2.449	2.738	3.365	3.622
34		1.691	2.032	2.441	2.728	3.348	3.601
36		1.688	2.028	2.434	2.719	3.333	3.582
38		1.686	2.024	2.429	2.712	3.319	3.566
40		1.684	2.021	2.423	2.704	3.307	3.551
42		1.682	2.018	2.418	2.698	3.296	3.538
44		1.680	2.015	2.414	2.692	3.286	3.526
46		1.679	2.013	2.410	2.687	3.277	3.515
48		1.677	2.011	2.407	2.682	3.269	3.505
50		1.676	2.009	2.403	2.678	3.261	3.496
60		1.671	2.000	2.390	2.660	3.232	3.460
70		1.667	1.994	2.381	2.648	3.211	3.435
80		1.664	1.990	2.374	2.639	3.195	3.416
90		1.662	1.987	2.368	2.632	3.183	3.402
100		1.660	1.984	2.364	2.626	3.174	3.390
120		1.658	1.980	2.358	2.617	3.160	3.373
150		1.655	1.976	2.351	2.609	3.145	3.357
200		1.653	1.972	2.345	2.601	3.131	3.340
300		1.650	1.968	2.339	2.592	3.118	3.323
400		1.649	1.966	2.336	2.588	3.111	3.315
500		1.648	1.965	2.334	2.586	3.107	3.310
600		1.647	1.964	2.333	2.584	3.104	3.307
∞		1.645	1.960	2.326	2.576	3.090	3.291