

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



Bachelor of Industrial Studies Honors /  
Bachelor of Software Engineering Honors

**Final Examination (2021/2022)**  
**MHZ4357: Applied Statistics**

**Index No:** .....

**Date:** 20/02/2023

**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, which is represented by the following stem-and-leaf diagram?

0		1
1		2 2
2		1 5
3		2 4 4 7
4		1 3 5 8 5
5		2 3 7 3 5 6 6 8 9
6		2 3 4 7 8 7 7 8 8 9 9 9
7		2 6 6

Answer:.....

.....

(Marks 10)

3. In a statistical analysis it was found that the null hypothesis of a test was accepted at 5% significant level. If the same statistical analysis is done at 1% significant level, then explain about the result of the test.

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

4. An quality assurance officer inspects a box of 5 tomatoes until they found a defective tomato. Write down the sample space for this quality assurance inspection.

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

5. Let  $A$  and  $B$  be two mutually exclusive events such that  $P(A) = \frac{1}{5}$  while  $P(A \cup B) = \frac{1}{2}$ . What is the value of  $P(B)$ ?

Answer:..... (Marks 10)

6. In a study of a random sample of 100 coconut trees to identify whether there is any relationship between their harvested coconuts and their age, the correlation coefficient is given as  $r = -0.03$ . Explain the meaning of the value of  $r$  here.

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

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

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

8. 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 margin of error in this calculation.

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

9. The 99% confidence interval for the mean length of frog jumps is given as  $(12.64, 14.44)cm$ . Interpret this confidence interval in a meaningful way.

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

10. A software engineer claims that the average number of faults obtain in a certain software development is more than 4. Write down the corresponding null and alternative hypothesis for the above claim.

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

### Part B

1. (a) The following table provide the data regarding number of faults with their sizes received from two different software development teams within the three month period.

Size of the fault	Number of Faults	
	Team A	Team B
Small	36	44
Medium	72	65
Large	24	15

Draw an appropriate graph to present the above data and interpret your graph in appropriate way. (Marks 25)

- (b) The expected hours of completing a certain task in the process of software tool development for a sample of 15 software tools are given in the following data list. Here it is assumed that each of these software tool are having same kind of characteristics until they complete the given task.

81	84	85	87	87
87	90	90	91	92
92	94	95	95	95

- i. Compile a table showing the frequency distribution and the relative frequency distribution where it is given that the number of classes is 5. (Marks 30)
  - ii. Draw the histogram for the above frequency distribution. (Marks 25)
  - iii. Draw the frequency polygon for above frequency distribution on the same diagram of the part ii. (Marks 10)
  - iv. Comment about the shape of the distribution of the sample data. (Marks 10)
2. A block of 29 tomatoes were sent for a quality test and provide a value for the quality in a scale from 0 to 50. The findings results are presented using an ordered stem-and-leaf diagram.

0	2
1	6
2	1, 6, 8, 8
3	3, 3, 3, 4, 5, 7, 9, 9
4	1, 2, 3, 4, 4, 5, 6, 9, 9
5	0, 0, 0, 0, 0

- (a) Determine the central tendency measures (Mean, Median and Mode) of the above sample. (Marks 20)
- (b) What are the best suited measures for the central tendency. Explain your answer. (Marks 10)
- (c) Find the range and inter-quartile range for the sample data. (Marks 30)
- (d) Draw the box-whisker plot for the above data. (Marks 20)
- (e) Are there any outliers in the given data set? Justify your answer. (Marks 20)
3. In an experiment, it was found that the number of failures occurs in a process of developing new mobile application, follows the probability mass function given in the following table.
- Let  $X$  be the random variable which represent the number of failures in the developing process of new mobile application.
- |            |     |     |     |     |     |           |
|------------|-----|-----|-----|-----|-----|-----------|
| $X = x$    | 1   | 2   | 3   | 4   | 5   | Otherwise |
| $P(X = x)$ | 0.1 | 0.2 | 0.2 | 0.3 | 0.2 | 0         |
- (a) Identify the set of all possible outcomes that the random variable  $X$  can has. (Marks 10)
- (b) Find the probability that it can be having more than three failures within the developing process of new mobile application. (Marks 20)
- (c) Find the probability that it can be having more than two failures and less than or equal four failures within the developing process of new mobile application. (Marks 30)
- (d) Find the expected number of failures it can be occur in the developing process of new mobile application. (Marks 40)
4. Every morning, Saman plans to catch the train in order to transport his farm's vegetables to a certain buyer. The train is scheduled to depart at 6:30 AM. The number of minutes after 6:30 AM in which the train departs can be modelled by the random variable  $X$ , which has a continuous uniform distribution over the interval  $[\alpha, \beta]$ , where  $0 \leq \alpha < \beta$ .
- (a) If the mean and the variance of the random variable  $X$  is 4 and 3, show that  $\alpha = 1$  and  $\beta = 7$ . (Hint:  $E[X] = \frac{\alpha+\beta}{2}$  and  $V[X] = \frac{(\beta-\alpha)^2}{12}$ ). (Marks 30)
- (b) Obtain the probability that Saman will catch the train if he is 5 minutes late at 6.30 AM. (Marks 20)

- (c) The probability that Saman completes the packing on time each morning is, 0.96. If Saman does not complete packing on time, he will definitely lose the buyer. If he completes the packing on time, he will catch the train on time, but he will lose the buyer if the train departs after 6.35 AM.
- Find the probability that Saman will lose the buyer. *(Marks 30)*
  - Given that Saman loses the buyer, determine the probability that he did not completed the packing on time. *(Marks 20)*
5. (a) Define the following terms.
- Interval estimate *(Marks 10)*
  - Standard error of the statistic *(Marks 10)*
- (b) A software development company has developed a new tool for banking mobile app which having new features. In a survey which was done at the introductory session of the new app, it was observed that 40% of the bankers are willing to purchase this app within next six month. 100 banking companies were participated for this introductory session.
- Give an estimate for the true proportion of bankers that are likely to purchase new banking app over the next six month period. *(Marks 10)*
  - Find the standard error of the true proportion of bankers that are likely to purchase new banking app over the next six month period. *(Marks 15)*
  - If the software company wants to find a 90% confidence interval for the true proportion of bankers that are likely to purchase new banking app over the next six month period, then
    - Find the margin of error for this calculation. *(Marks 20)*
    - Find the confidence Interval and interpret it in a meaningful way. *(Marks 35)*
6. An electronic machine is programmed to packet the 5kg rice packets. However, rice company owner had complained to the programmer that the weight of the rice packets which are packed by the said machine are greater 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 packed by the said machine, and identified that the sample mean weight of rice is 5.2kg and sample variance is 0.25kg per packet. Then the programmer use hypothesis testing to see the validity of the rice company owners claim.

- (a) State the corresponding null and alternative hypotheses that the programmer of the rice packeting machine test to see the validity of the rice company owners 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 rice company owner. (Marks 35)
7. An owner of a tea estate interested in the correlation between the harvested tea amount(Kg) per plant and the age(years) of the plant used. The data gathered from a sample of 20 tea plants and data was used to determine whether the model follows a simple linear regression model. Following R output were obtain within this study.

Call:

```
lm(formula = HarvestedTea ~ Age)
```

Coefficients:

(Intercept)	Age
26.2503	-0.3697

Call:

```
lm(formula = HarvestedTea ~ Age)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.3318	-0.5228	0.2883	0.6619	1.0600

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	26.25031	0.45327	57.91	< 2e-16 ***
Age	-0.36970	0.02965	-12.47	2.72e-10 ***

---

Signif. codes: 0 '\*\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9901 on 18 degrees of freedom

Multiple R-squared: 0.8962, Adjusted R-squared: 0.8904

F-statistic: 155.4 on 1 and 18 DF, p-value: 2.724e-10

- (a) Identify the estimated slop and intercept coefficients of the fitted model. Hence express the fitted model for harvested tea amount and age of the plant, then interpret the model. (Marks 25)
- (b) Discuss about the accuracy level of the fitted model. (Marks 15)
- (c) Following table represent the ANOVA which is used to test the accuracy of the model coefficients.

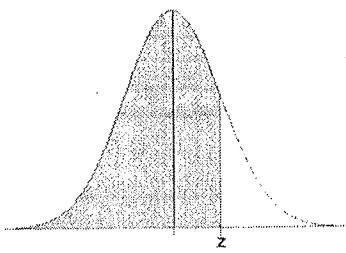
Source	Df	Sum Sq	Mean Sq	F value	P-value
Age	A	152.345	B	C	2.724e-10 ***
Residuals	18	17.644	0.98		
Total	19	D			

- i. State the null and alternative hypothesis for this ANOVA. (Marks 10)
- ii. Find the values of A, B, C and D. (Marks 20)
- iii. At 5% significant level discuss the conclusion of the ANOVA. (Marks 20)
- (d) State the assumptions that you have to make within this analysis. (Marks 10)

End.

Copyright Reserved.

## Standard Normal Cumulative Probability Table



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

*F* Distribution: Critical Values of *F* (5% significance level)

<i>v<sub>1</sub></i>	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20
<i>v<sub>2</sub></i>															
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	239.88	240.54	241.88	243.91	245.36	246.46	247.32	248.01
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.42	19.43	19.44	19.45
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.71	8.69	8.67	8.66
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.87	5.84	5.82	5.80
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.64	4.60	4.58	4.56
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.96	3.92	3.90	3.87
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.53	3.49	3.47	3.44
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.24	3.20	3.17	3.15
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.03	2.99	2.96	2.94
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.86	2.83	2.80	2.77
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.74	2.70	2.67	2.65
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.64	2.60	2.57	2.54
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.55	2.51	2.48	2.46
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.48	2.44	2.41	2.39
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.42	2.38	2.35	2.33
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.37	2.33	2.30	2.28
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.33	2.29	2.26	2.23
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.29	2.25	2.22	2.19
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.26	2.21	2.18	2.16
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.22	2.18	2.15	2.12
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.20	2.16	2.12	2.10
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.17	2.13	2.10	2.07
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.15	2.11	2.08	2.05
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.13	2.09	2.05	2.03
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.11	2.07	2.04	2.01
26	4.22	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.09	2.05	2.02	1.99
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.08	2.04	2.00	1.97
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.06	2.02	1.99	1.96
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.05	2.01	1.97	1.94
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.04	1.99	1.96	1.93
35	4.12	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.16	2.11	2.04	1.99	1.94	1.91	1.88
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.95	1.90	1.87	1.84
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.95	1.89	1.85	1.81	1.78
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.86	1.82	1.78	1.75
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02	1.97	1.89	1.84	1.79	1.75	1.72
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95	1.88	1.82	1.77	1.73	1.70
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94	1.86	1.80	1.76	1.72	1.69
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.85	1.79	1.75	1.71	1.68
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.78	1.73	1.69	1.66
150	3.90	3.06	2.66	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.82	1.76	1.71	1.67	1.64
200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.80	1.74	1.69	1.66	1.62
250	3.88	3.03	2.64	2.41	2.25	2.13	2.05	1.98	1.92	1.87	1.79	1.73	1.68	1.65	1.61
300	3.87	3.03	2.63	2.40	2.24	2.13	2.04	1.97	1.91	1.86	1.78	1.72	1.68	1.64	1.61
400	3.86	3.02	2.63	2.39	2.24	2.12	2.03	1.96	1.90	1.85	1.78	1.72	1.67	1.63	1.60
500	3.86	3.01	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.77	1.71	1.66	1.62	1.59
600	3.86	3.01	2.62	2.39	2.23	2.11	2.03	1.95	1.90	1.85	1.77	1.71	1.66	1.62	1.59
750	3.85	3.01	2.62	2.38	2.23	2.11	2.03	1.95	1.89	1.84	1.77	1.70	1.66	1.62	1.58
1000	3.85	3.00	2.61	2.38	2.22	2.11	2.02	1.95	1.89	1.84	1.76	1.70	1.65	1.61	1.58

F Distribution: Critical Values of F (5% significance level)

$v_1$	25	30	35	40	50	60	75	100	150	200
$v_2$										
1	249.36	250.10	250.69	251.14	251.77	252.20	252.62	253.04	253.46	253.68
2	19.46	19.46	19.47	19.47	19.48	19.48	19.48	19.49	19.49	19.49
3	8.63	8.62	8.60	8.59	8.58	8.57	8.56	8.55	8.54	8.54
4	5.77	5.75	5.73	5.72	5.70	5.69	5.68	5.66	5.65	5.65
5	4.52	4.50	4.48	4.46	4.44	4.43	4.42	4.41	4.39	4.39
6	3.83	3.81	3.79	3.77	3.75	3.74	3.73	3.71	3.70	3.69
7	3.40	3.38	3.36	3.34	3.32	3.30	3.29	3.27	3.26	3.25
8	3.11	3.08	3.06	3.04	3.02	3.01	2.99	2.97	2.96	2.95
9	2.89	2.86	2.84	2.83	2.80	2.79	2.77	2.76	2.74	2.73
10	2.73	2.70	2.68	2.66	2.64	2.62	2.60	2.59	2.57	2.56
11	2.60	2.57	2.55	2.53	2.51	2.49	2.47	2.46	2.44	2.43
12	2.50	2.47	2.44	2.43	2.40	2.38	2.37	2.35	2.33	2.32
13	2.41	2.38	2.36	2.34	2.31	2.30	2.28	2.26	2.24	2.23
14	2.34	2.31	2.28	2.27	2.24	2.22	2.21	2.19	2.17	2.16
15	2.28	2.25	2.22	2.20	2.18	2.16	2.14	2.12	2.10	2.10
16	2.23	2.19	2.17	2.15	2.12	2.11	2.09	2.07	2.05	2.04
17	2.18	2.15	2.12	2.10	2.08	2.06	2.04	2.02	2.00	1.99
18	2.14	2.11	2.08	2.06	2.04	2.02	2.00	1.98	1.96	1.95
19	2.11	2.07	2.05	2.03	2.00	1.98	1.96	1.94	1.92	1.91
20	2.07	2.04	2.01	1.99	1.97	1.95	1.93	1.91	1.89	1.88
21	2.05	2.01	1.98	1.96	1.94	1.92	1.90	1.88	1.86	1.84
22	2.02	1.98	1.96	1.94	1.91	1.89	1.87	1.85	1.83	1.82
23	2.00	1.96	1.93	1.91	1.88	1.86	1.84	1.82	1.80	1.79
24	1.97	1.94	1.91	1.89	1.86	1.84	1.82	1.80	1.78	1.77
25	1.96	1.92	1.89	1.87	1.84	1.82	1.80	1.78	1.76	1.75
26	1.94	1.90	1.87	1.85	1.82	1.80	1.78	1.76	1.74	1.73
27	1.92	1.88	1.86	1.84	1.81	1.79	1.76	1.74	1.72	1.71
28	1.91	1.87	1.84	1.82	1.79	1.77	1.75	1.73	1.70	1.69
29	1.89	1.85	1.83	1.81	1.77	1.75	1.73	1.71	1.69	1.67
30	1.88	1.84	1.81	1.79	1.76	1.74	1.72	1.70	1.67	1.66
35	1.82	1.79	1.76	1.74	1.70	1.68	1.66	1.63	1.61	1.60
40	1.78	1.74	1.72	1.69	1.66	1.64	1.61	1.59	1.56	1.55
50	1.73	1.69	1.66	1.63	1.60	1.58	1.55	1.52	1.50	1.48
60	1.69	1.65	1.62	1.59	1.56	1.53	1.51	1.48	1.45	1.44
70	1.66	1.62	1.59	1.57	1.53	1.50	1.48	1.45	1.42	1.40
80	1.64	1.60	1.57	1.54	1.51	1.48	1.45	1.43	1.39	1.38
90	1.63	1.59	1.55	1.53	1.49	1.46	1.44	1.41	1.38	1.36
100	1.62	1.57	1.54	1.52	1.48	1.45	1.42	1.39	1.36	1.34
120	1.60	1.55	1.52	1.50	1.46	1.43	1.40	1.37	1.33	1.32
150	1.58	1.54	1.50	1.48	1.44	1.41	1.38	1.34	1.31	1.29
200	1.56	1.52	1.48	1.46	1.41	1.39	1.35	1.32	1.28	1.26
250	1.55	1.50	1.47	1.44	1.40	1.37	1.34	1.31	1.27	1.25
300	1.54	1.50	1.46	1.43	1.39	1.36	1.33	1.30	1.26	1.23
400	1.53	1.49	1.45	1.42	1.38	1.35	1.32	1.28	1.24	1.22
500	1.53	1.48	1.45	1.42	1.38	1.35	1.31	1.28	1.23	1.21
600	1.52	1.48	1.44	1.41	1.37	1.34	1.31	1.27	1.23	1.20
750	1.52	1.47	1.44	1.41	1.37	1.34	1.30	1.26	1.22	1.20
1000	1.52	1.47	1.43	1.41	1.36	1.33	1.30	1.26	1.22	1.19