00297

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

Bachelor of Software Engineering Honours

M

Final Examination (2023/2024)
MHZ4377: Applied Statistics

Date: 24th January 2025 (Friday) **Time: 0930 hrs. – 1230 hrs.**

Instructions:

Index. No:

- ***Part A is Compulsory***
  - *Provide short answers in the given space.*
  - *Do not need to show any workings.*
- *Answer **five (05)** questions only out of seven (07) questions from **Part B**.*
  - *Provide answers in separate sheets (answer booklet) which will be provided in the examination.*
  - *Show all your workings.*
- *The number of pages in the paper is **Six (06)**.*
- *All the symbols are in standard notation unless they are defined.*
- *All the relevant statistical tables and equations are provided.*
- ***Attach Part A to the Answer script of Part B.***
- ***Non-Programmable Calculators are allowed.***

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. In a workshop the lecturer wants to create pairs among the group consisting of 2 engineers and 3 designers. Two people are selected randomly, one at a time, without replacement. Write down the sample space for this experiment.

Answer.....

[10%]

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

Answer.....

[10%]

3. Suppose a survey is conducted in a city with a sample of 400 people, and 180 of them said they prefer a particular brand of soda. Calculate the standard error for the sample proportion who prefer the said brand of soda.

Answer.....

[10%]

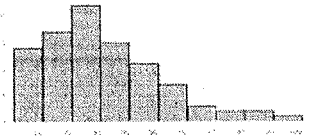
4. In a study of a random sample of 100 coconut trees to identify whether there is any relationship between their harvested coconuts and their age, the correlation coefficient is given as r = -0.3. Explain the meaning of the value of r here.

Answer.....

[10%]

5. What can you say about the shape of the distribution, which is represented by the histogram?

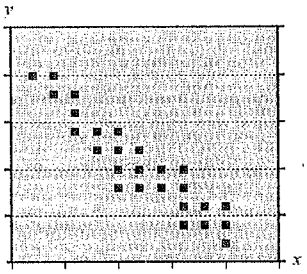
[10%]



Answer.....
.....
.....

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

[10%]



Answer.....
.....

7. A software engineer claims that the average time required to process a transaction in a software system is less than 3 seconds. Write down the corresponding null and alternative hypothesis for the above claim.

Answer.....

[10%]

8. If the successive probability of a Bernoulli trail is given as 0.2, what is the expected value of the random variable X which follows the Geometric distribution with the above Bernoulli trail?

Answer.....

[10%]

9. Given the sample mean of variable X is $\bar{X} = 180$. Suppose the random variable Y having the linear relationship with X as, $Y = 2X - 100$, then find the sample mean of variable Y.

Answer.....

[10%]

10. Given the observations $x, 5, 6, y, 7, 8$ in ascending order. Suppose that both the mean and the mode of these observations are 6. What are the values of x and y?

Answer.....

[10%]

Part B

Q1. The following are the response times (in milliseconds) recorded for 25 API calls during system performance testing:

125	128	130	132	135	138	140	142	144	146
148	150	150	152	154	156	158	160	162	164
166	170	175	180	184					

(a) Construct a group frequency distribution table with five classes. **[30%]**

(b) Draw the histogram for the above frequency distribution. **[25%]**

(c) Draw the frequency polygon for the above distribution on the same diagram of part (b). **[25%]**

(d) Comment about the shape of the distribution of the data. **[10%]**

(e) Based on the results, discuss whether the response times are evenly distributed or if there are bottlenecks in the system. **[10%]**

Q2. The time (minutes) it needs to do a certain computer game by 20 players were coded by subtracting 10 from each of times and then dividing by 2. The ungroup frequency distribution of the coded data with two missing frequencies is given below.

Coded times (in Minutes)	Frequencies
05	2
10	f_1
15	9
20	f_2
25	1

The estimated mean for the coded times is given as 14 minutes.

(a) Find the values of f_1 and f_2. **[20%]**

(b) Find the median and mode of the coded times. **[20%]**

(c) Find the variance and interquartile range for coded time. **[25%]**

(d) Hence estimate the central tendency measures and dispersion measures of the actual times taken to finish this computer game. **[25%]**

(e) Comment about the shape of the distribution of the data. **[10%]**

Q3. Suppose that the time (minutes) it needs to run a certain computer programme is given by the random variable T, with the probability density function:

$$f_T(t) = \begin{cases} k - \frac{1}{2}t; & 0 \leq t \leq 2 \\ 0 & otherwise \end{cases}$$

Here $k \in \mathbb{R}$.

(a) Find the value of k. **[20%]**

(b) Find the mean running time of the programme and the variance of the running time of a certain programme. **[50%]**

(c) Find the cumulative distribution function of the random variable running time of the programme. **[20%]**

(d) What is the probability that the running time of the programme is more than one minute? **[10%]**

Q4.

(a) Discuss the characteristics of Geometric distribution and the characteristic of Binomial distribution. **[20%]**

(b) A worn machine is known to produce 10% defective components. If the random variable X is the number of defective components produced in a run of 40 components,

(i) Calculate the probability that two defective items are produced. **[20%]**

(ii) Calculate the probability that more than two defective items are produced. **[30%]**

(iii) Find the average number of defectives among run of 40 components and the variance of the distribution of defective items among run of 40 components. **[30%]**

Q5. A software company is conducting performance testing for a new application to measure the average time (in seconds) it takes for a user to complete a specific task within the application. The company collected data from a random sample of 20 users. The recorded average time is 15.6 seconds, and the standard deviation of the recorded time is 3.2 seconds.

(a) Find the point estimate of the time taken to complete said task. **[10%]**

(b) Find the standard error of the sample mean of the recorded time. **[15%]**

(c) If the software company wants to find a 90% confidence interval for the recorded time, then

(i) Find the margin of error for this calculation. **[20%]**

(ii) Find the confidence interval and interpret it in a meaningful way. **[30%]**

(d) Find the probability that true average time needed to complete this task is between 14.10 second to 17. 10 seconds. **[25%]**

Q6. An automated system is used by a certain auto mobile service station where they are sending automated messages for their regular customers to remind service due date based on the distance there were travel by their vehicle within 6-month period. Whenever the distance travel by a given vehicle is greater than 1500 Km, an automated message will be sent to the customer. However, some of the customers were saying that they did not receive these messages on time. To test this claim about the automated system, the software engineer collects the details travel by each vehicle within every 6 months from 36 vehicles who are regular customers in this service center with no message received by the service station. According to this data it was found that the average distance travel in every 6-month period is 1570 Km and standard deviation of the distance travel in 6-month period is 150.

(a) State the corresponding null hypotheses and alternative hypothesis to be tested. Comment whether this is two tail test or one tail test. **[25%]**

(b) Find the test statistics associated with this hypothesis testing. **[25%]**

(c) If the analyst wants to test this hypothesis at 5% significant level, then what is the critical value/s associated with this test? Explain why you chose the corresponding table when finding critical value/s **[20%]**

(d) Test the hypothesis at 5% significant level and state your decision regarding the hypothesis. Explain your conclusion regarding the claim of the customers. **[30%]**

Q7. A software company wants to examine whether the number of hours of training a software engineer receives has an impact on their performance rating. A sample of 12 software engineers was taken, and the following data was collected:

Hours	5	8	12	6	4	10	15	7	9	3	11	14
Performance Rating (out of 10)	6	7	8	6	5	9	9	6	8	5	8	9

(a) Identify the independent variable and the dependent variable. **[10%]**

(b) Calculate the correlation coefficient for above data. Interpret your answer. **[30%]**

(c) Draw the scatter plot for dependent variable against the independent variable. **[10%]**

(d) Find the coefficient of the equation of the best fitted simple linear regression model of the above data. **[30%]**

(e) Interpret the found regression coefficient in part (d). **[10%]**

(f) Use the estimated line of regression to find the performance rate of a software engineer who has 12-hour training. **[10%]**

End.

Important Formulars:

- Binomial distribution

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

- Geometric distribution

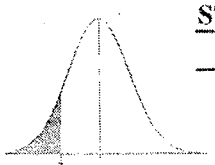
  $$Pr(X = r) = pq^{r-1}$$

- Poison Distribution

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

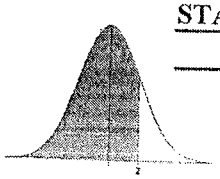
- If random variable X, follows normal distribution ($X \sim N(\mu, \sigma^2)$), Then $Z = \frac{X-\mu}{\sigma}$ follows standard normal distribution where $Z \sim N(1,0)$.
- $r = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}$
- $\beta_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$

Standard Normal Distribution Tables



STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00004	.00004	.00003	.00003
-3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00006	.00005	.00005	.00005
-3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00008	.00008	.00008	.00008
-3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013	.00012	.00012	.00011
-3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019	.00018	.00017	.00017
-3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025	.00024
-3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039	.00038	.00036	.00035
-3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056	.00054	.00052	.00050
-3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00079	.00076	.00074	.00071
-3.0	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104	.00100
-2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00154	.00149	.00144	.00139
-2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212	.00205	.00199	.00193
-2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289	.00280	.00272	.00264
-2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00391	.00379	.00368	.00357
-2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00523	.00508	.00494	.00480
-2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639
-2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00914	.00889	.00866	.00842
-2.2	.01390	.01355	.01321	.01287	.01255	.01222	.01191	.01160	.01130	.01101
-2.1	.01786	.01743	.01700	.01659	.01618	.01578	.01539	.01500	.01463	.01426
-2.0	.02275	.02222	.02169	.02118	.02068	.02018	.01970	.01923	.01876	.01831
-1.9	.02872	.02807	.02743	.02680	.02619	.02559	.02500	.02442	.02385	.02330
-1.8	.03593	.03515	.03438	.03362	.03288	.03216	.03144	.03074	.03005	.02938
-1.7	.04457	.04363	.04272	.04182	.04093	.04006	.03920	.03836	.03754	.03673
-1.6	.05480	.05370	.05262	.05155	.05050	.04947	.04846	.04746	.04648	.04551
-1.5	.06681	.06552	.06426	.06301	.06178	.06057	.05938	.05821	.05705	.05592
-1.4	.08076	.07927	.07780	.07636	.07493	.07353	.07215	.07078	.06944	.06811
-1.3	.09680	.09510	.09342	.09176	.09012	.08851	.08691	.08534	.08379	.08226
-1.2	.11507	.11314	.11123	.10935	.10749	.10565	.10383	.10204	.10027	.09853
-1.1	.13567	.13350	.13136	.12924	.12714	.12507	.12302	.12100	.11900	.11702
-1.0	.15866	.15625	.15386	.15151	.14917	.14686	.14457	.14231	.14007	.13786
-0.9	.18406	.18141	.17879	.17619	.17361	.17106	.16853	.16602	.16354	.16109
-0.8	.21186	.20897	.20611	.20327	.20045	.19766	.19489	.19215	.18943	.18673
-0.7	.24196	.23885	.23576	.23270	.22965	.22663	.22363	.22065	.21770	.21476
-0.6	.27425	.27093	.26763	.26435	.26109	.25785	.25463	.25143	.24825	.24510
-0.5	.30854	.30503	.30153	.29806	.29460	.29116	.28774	.28434	.28096	.27760
-0.4	.34458	.34090	.33724	.33360	.32997	.32636	.32276	.31918	.31561	.31207
-0.3	.38209	.37828	.37448	.37070	.36693	.36317	.35942	.35569	.35197	.34827
-0.2	.42074	.41683	.41294	.40905	.40517	.40129	.39743	.39358	.38974	.38591
-0.1	.46017	.45620	.45224	.44828	.44433	.44038	.43644	.43251	.42858	.42465
-0.0	.50000	.49601	.49202	.48803	.48405	.48006	.47608	.47210	.46812	.46414



STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
3.0	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983
3.6	.99984	.99985	.99985	.99986	.99986	.99987	.99987	.99988	.99988	.99989
3.7	.99989	.99990	.99990	.99990	.99991	.99991	.99992	.99992	.99992	.99992
3.8	.99993	.99993	.99993	.99994	.99994	.99994	.99994	.99995	.99995	.99995
3.9	.99995	.99995	.99996	.99996	.99996	.99996	.99996	.99996	.99997	.99997

t Distribution: Critical Values of *t*

Degrees of freedom	*Two-tailed test:*	*Significance level* 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