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(2019/2020)

MHZ4357: Applied Statistics

Index No:.....

Date: 30^{th} July 2020 (Thursday) Time: 13:30 hrs. – 16:30 hrs.

Instructions:

- Part A is compulsory.
  - Provide short answers in given space.
  - Don't have to show any working
- Answer any FIVE(5) questions form Part B.
  - Provide answers in a separate answer scripts given by the university.
  - State any assumptions you required.
  - Show all your workings.
- All symbols are in standard notation.
- This paper contains TEN(10) pages.
- Attached the Part A to the Answer Scrip of the Part B.

Part A

Provide short answers in given space. You don't have to show any work. Each answer is worth 10 marks.

1. In 3 tosses of a coin which of following equals the event "exactly two heads"?
   $A = \{THH; HTH; HHT; HHH\}$
   $B = \{THH; HTH; HHT\}$
   $C = \{THH; HHT\}$

   Answer .....

2. Let P and Q be two events defined on the same sample space S. Suppose that the probability that neither event occurs is 3/8. What is the probability that at least one of the events occurs?

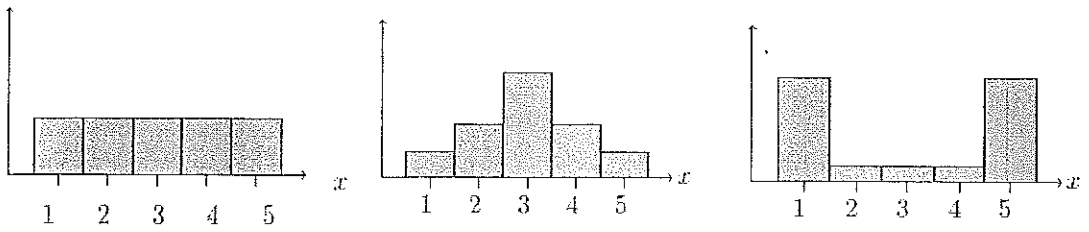
   Answer:.....

3. In testing a hypothesis,suppose you got a p-value of 0.184. Given 10% level of significance what is your conclusion about the hypothesis?

   Answer:.....

4. A 95% confidence interval for the mean number of televisions per Sri Lankan household is (1.15, 4.20). What is the meaning of this statement?

   Answer:.....



(a) A (b) B (c) C

Figure 1: The probability mass functions for three different random variables

5. The graphs in Figure 1, give the probability mass functions for three different random variables. Order the graphs from smallest to biggest standard deviation.

   Answer .....

Consider the following scenario for Questions (6) and (7)

A Hypothesis testing was done to examine whether the mean cholesterol concentration in two groups: Control Group and the Exercise Group are the same or not. There were 10 participants in the control group and 11 participants in the exercise group. Figure 2 represent the R-Output for the relevant hypothesis test.

```
> t.test(cholesterol ~ group, var.equal=TRUE, data = istt)

        Two Sample t-test

data:  cholesterol by group
t = 3.776, df = 19, p-value = 0.001278
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.2274728 0.7932545
sample estimates:
 mean in group control mean in group exercise
              5.064000                4.553636
```

Figure 2: Test Results of the R Output

6. Write down the null and alternative hypothesis for this problem.

Answer .....

.....

7. Interpret the test output given in Figure 2.

Answer:.....

State whether the following statements given in Questions (8) and (9), regarding the sample distribution of $\bar{x}$ are true or false.

8. If you increase your sample size, then the mean of the sample distribution of $\bar{x}$ will always get closer to the population mean μ,.

Answer .....

9. The standard deviation of the sample distribution of the sample mean($\bar{x}$) is same as the standard deviation of the population.

Answer .....

10. What is the goal of the descriptive statistics?

Answer:.....

Part B

1. (a) Define following terms.

   i. Population

   (5 marks)

   ii. Sample

   (5 marks)

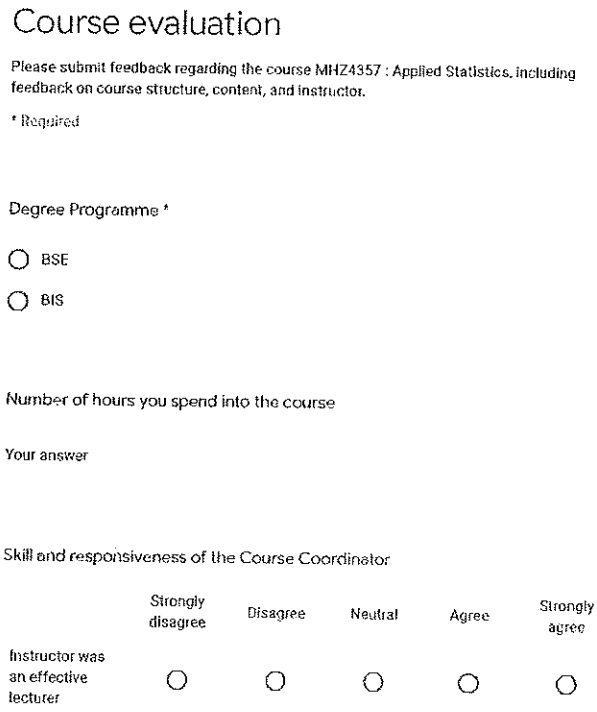
   iii. Parameter

   (5 marks)

   iv. Statistic

   (5 marks)

   (b) Consider the following questionnaire given in Figure 3.



Course evaluation

Please submit feedback regarding the course MHZ4357 : Applied Statistics, including feedback on course structure, content, and instructor.

* Required

Degree Programme *

BSE

BIS

Number of hours you spend into the course

Your answer

Skill and responsiveness of the Course Coordinator

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Instructor was an effective lecturer					

Figure 3: A sample Questionnaire

   i. What is the scale of measurement used for the item "Degree Programme"?

   (5 marks)

   ii. What is the type of data collected(or referred) in the item "Skill and responsiveness of the Course Coordinator"?

   (5 marks)

iii. Give an example of a variable mentioned in the questionnaire (Figure 3) which is in the ratio scale.

(5 marks)

(c) Table 1 provides the data obtained for the item "Instructor was an effective lecturer" for a random sample of 40 learners.

Table 1: The Data regarding the variable "Instructor was an effective lecturer"

1	1	2	2	4	4	5	1	2	1
2	5	4	3	5	2	2	3	5	2
5	1	1	2	1	3	1	2	3	4
5	1	5	4	3	5	4	4	5	1

Where the codes represent,
1:Strongly Disagree
2:Disagree
3:Neutral
4:Agree
5:Strongly Agree

i. Summarize the data given in Table 1, in a form of frequency table which include the absoluate frequencies, cumulative frequencies and relative frequencies.

(15 marks)

ii. Interpret the frequency table that you have created in part 1(c)i .

(10 marks)

(d) Table 2 represents the summary of the data obtained for the variable "Number of hours spent per week for studying MHZ4357 " for a sample of 50 learners

i. Draw the corresponding histogram for the data given in Table 2 .

(20 marks)

ii. Comment on the shape of the histogram.

(10 marks)

iii. Draw the Ogive for the histogram in part 1(d)i

(10 marks)

Table 2: The frequency table for the variable
"Number of hours spend per week for studying MHZ4357"

Class	Frequencies
$0 < x \leq 5$	2
$6 < x \leq 15$	9
$16 < x \leq 20$	8
$21 < x \leq 25$	14
$26 < x \leq 30$	9
$31 < x \leq 35$	6
$36 < x \leq 40$	2

2. Consider the following data about the life time (in Hours) of a certain electrical component of a random sample.

195 , 90 , 92 , 89 , 93 , 91 , 88 , 89 , 94 , 95

(a) Find the central tendency measures(Mean, Médian and Mode) of the life time (in Hours) of the given electrical component.

(20 marks)

(b) Find range, standard deviation and inter-quartile range of the life time (in Hours) of the given electrical component.

(30 marks)

(c) Comment on the shape of the distribution of the life time (in Hours) of the given electrical component.

(10 marks)

(d) What is the best suited measure for central tendency? Explain.

(10 marks)

(e) Are there any outliers in this data set? Explain.

(10 marks)

(f) If outliers are present in the data set, suppose they are removed and measures of central tendency and variance are recalculated. Comment on the shape of distribution of the new data set.

(20 marks)

3. (a) Define following terms.

i. Sample Space

(10 marks)

ii. An event

(10 marks)

(b) i. Let A and B be two events in a given sample space. Suppose that the probability that both events occurs is $\frac{1}{8}$ and neither event occurs is $\frac{3}{8}$. What is the probability that exactly one of the event occur?

(20 marks)

ii. Let C and D be two events in a given sample space. Suppose $P(C) = 0.5$, $P(C \cap D) = 0.2$ and $P((C \cup D)^c) = 0.4$. What is $P(D)$?

(20 marks)

(c) Three different machines: $M1, M2$ and $M3$ are used to produce similar electric components. Machines $M1, M2$ and $M3$ produced 20%, 30% and 50% of the components respectively. It is known the probabilities that the machines produce defective components are 1% for $M1$, 2% for $M2$ and 3% for $M3$.

i. If a component is selected randomly from a large batch and that component is defective. Find the probability that was produced by $M3$.

(20 marks)

ii. If a component is selected randomly from a large batch and what is the probability that component will be a defective?

(20 marks)

4. (a) A continuous random variable X represents the life time of a machine part (in months). The probability density function for random variable X is defined on the interval (0,36) as,

$$f_X(x) = \frac{1}{12\sqrt{36 - x}} \qquad \text{where} \quad 0 \leq x \leq 36$$

i. Find the cumulative distribution function($F_X(x)$) of the random variable X. [*Hint:* $\int_p^q (ax + b)^n dx = \frac{1}{a} \frac{(ax+b)^{n+1}}{n+1} \Big|_p^q$.

(25 marks)

ii. Find the probability that the life time of randomly selected machine part will lie between 20 to 27 weeks .

(20 marks)

(b) On the basis of past experience, the probability that a certain electrical component will be satisfactory is 98%. The components are sampled item by item from continuous production. The satisfaction of any component is independent from that of the other component. If a sample of four components are selected at random, and the random variable X is defined as the number of defective components.

i. Find the probability that none of the above components are defective. .

(15 marks)

ii. Find the probability that at least one component is defective.

(10 marks)

iii. Find the expected number of defectives.

(15 marks)

iv. Find the variance of the random variable X: number of defectives.

(15 marks)

5. (a) Discuss the difference about the following terms.

i. Estimator Vs Estimate

(10 marks)

ii. Point estimate Vs Interval estimate

(10 marks)

(b) A researcher is interested about the mean viscosity of a product that produce in a batch process. Recorded below are the 12 most recent values, taken from consecutive batches.

13.7, 14.9, 15.7, 16.1, 14.7, 15.2, 13.9, 13.9, 15.0, 13.0, 16.7, 13.2

i. Find a point estimation for the population mean of the viscosity. Interpret your answer.

(20 marks)

ii. Find the sample standard deviation of the viscosity.

(15 marks)

iii. Find the standard error of the mean viscosity.

(10 marks)

iv. Find an interval estimate for the population mean at 5% significant level.

(25 marks)

v. Interpret your answer in part 5(b)iv.

(10 marks)

6. (a) State the properties of Null Hypothesis(H_0) and Alternative Hypothesis (H_1).

(10 marks)

(b) Discuss about the Type I error and Type II error about the hypothesis testing.

(10 marks)

(c) A firm produces steel wire of a particular gauge. The firm suspects that the mean breaking strength, in Newton(N)of wire is $82N$. The past studies suggested that the variance of the braking strength is $12n$. Below shows the braking strength of a random sample of wire.

80.5 83.1 73.6 70.4 68.9 71.6 82.3 78.6 73.4

i. Calculate sample mean of the breaking strength.

(10 marks)

ii. State the hypothesis for decision makers claim.

(20 marks)

iii. Test the above hypothesis using an appropriate method at 5% significant level.

(30 marks)

iv. Interpret your answer in part 6(c)iii.

(10 marks)

v. State the assumptions that you have made.

(10 marks)

7. The fasting blood glucose level mg/dl and the age t years of a random sample of 10 patients are shown in the table 3 below.

Table 3: The fasting blood glucose level mg/dl .

Patient	A	B	C	D	E	F	G	H	I	J
t	42	74	48	35	56	26	60	60	25	72
G	98	130	120	88	182	80	135	130	81	130

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

(b) Calculate the correlation coefficient for above data. (*25 marks*)

(c) Interpret the correlation coefficient. (*10 marks*)

(d) Draw the scatter Diagram for blood glucose level against age for these 10 patients. (*10 marks*)

(e) Find the equation for best fitted regression line of G on t . (*25 marks*)

(f) Draw the best fitted line on the scatter plot in part 7d. (*10 marks*)

(g) Use your regression line to estimate the blood pressure of a 40 years patient. (*10 marks*)

End.

Probabilities for the standard normal distribution

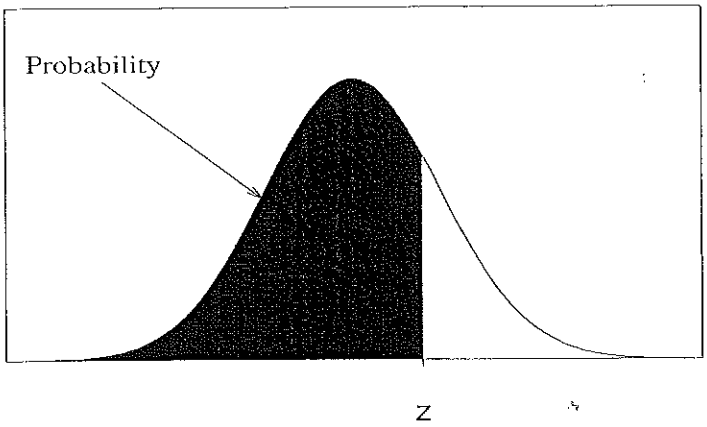
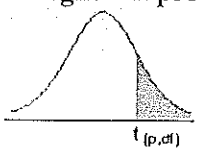


Table entry for z is the probability lying to the left of z

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.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 table with right tail probabilities



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	4.3178
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
inf	0.253347	0.674490	1.281552	1.644854	1.95996	2.32635	2.57583	3.2905

Notes

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A) = 1 - P(A')$$

Correlation coefficient Value

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Parameters of the linear regression model

$$\widehat{\beta_1} = \frac{\sum_{i=1}^{n}(x_i y_i) - \frac{(\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}}{\sum_{i=1}^{n} {x_i}^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}}$$

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

Best fitted line

$$\hat{y} = \widehat{\beta_0} - \widehat{\beta_1}x$$