UNIVERSITI TUN HUSSEIN ONN MALAYSIA

FINAL EXAMINATION
SEMESTER I
SESSION 2022/2023

COURSE NAME	:	ENGINEERING STATISTICS / ENGINEERING MATHEMATICS V
COURSE CODE	:	BEE32502 / BEE31702
PROGRAMME CODE	:	BEJ / BEV
EXAMINATION DATE	:	FEBRUARY 2023
DURATION	:	3 HOURS
INSTRUCTIONS	:	1. ANSWER **ALL** QUESTIONS IN THIS QUESTION PAPER. 2. THIS FINAL EXAMINATION IS CONDUCTED VIA **CLOSED BOOK**. 3. STUDENTS ARE **PROHIBITED** TO CONSULT THEIR OWN MATERIAL OR ANY EXTERNAL RESOURCES DURING THE EXAMINATION CONDUCTED VIA CLOSED BOOK

THIS QUESTION PAPER CONSISTS OF **EIGHTEEN (18)** PAGES

PART A: OBJECTIVE QUESTIONS

Q1 What does the central limit theorem state?

(a) if the sample size increases sampling distribution must approach normal distribution
(b) if the sample size decreases then the sample distribution must approach normal distribution
(c) if the sample size increases then the sampling distribution much approach an exponential distribution
(d) if the sample size decreases then the sampling distribution much approach an exponential distribution

(1 mark)

Q2 Which of the following is a subset of population?

(a) distribution
(b) sample
(c) data
(d) set

(1 mark)

Q3 The sampling error is defined as _________.

(a) difference between population and parameter
(b) difference between sample and parameter
(c) difference between population and sample
(d) difference between parameter and sample

(1 mark)

Q4 Which Chi Square distribution looks the most like a normal distribution?

(a) A Chi Square distribution with 4 degrees of freedom
(b) A Chi Square distribution with 5 degrees of freedom
(c) A Chi Square distribution with 6 degrees of freedom
(d) A Chi Square distribution with 16 degrees of freedom

(1 mark)

Q5 Which of the following distributions is continuous?

(a) Binomial Distribution
(b) Hyper-geometric Distribution
(c) F-Distribution
(d) Poisson Distribution

(1 mark)

Q6 An estimator is a random variable because it varies from ____________.

(a) Population to sample
(b) Population to population
(c) Sample to sample
(d) Sample to population

(1 mark)

Q7 If the population standard deviation σ is doubled, the width of the confidence interval for the population mean, μ (i.e.; the upper limit of the confidence interval – lower limit of the confidence interval) will be:

(a) Doubled
(b) Decreased
(c) Divided by 2
(d) Multiplied by $\sqrt{2}$

(1 mark)

Q8 A confidence interval will be widened if __________.

(a) The confidence level is increased, and the sample size is reduced
(b) The confidence level is increased, and the sample size is increased
(c) The confidence level is decreased, and the sample size is increased
(d) The confidence level is decreased, and the sample size is decreased

(1 mark)

Q9 Which of the following statements is incorrect?

(a) If the population mean and population standard deviation are both known, one can make probability statements about individual x values taken from the population
(b) If the population mean and population standard deviation are both known, one can use the central limit theorem and make probability statements about the means of samples taken from the population
(c) If the population mean is unknown, one can use sample data as the basis from which to make probability statements about the true (but unknown) value of the population mean
(d) When sample data are used for estimating a population mean, sampling error will not be present since the observed sample statistic will not differ from the actual value of the population parameter

(1 mark)

Q10 A 95% confidence interval for the population mean is calculated to be 75.29 to 81.45. If the confidence level is reduced to 90%, the confidence interval will __________.

(a) become narrower
(b) remain the same
(c) become wider
(d) double in size

(1 mark)

Q11 The average growth of a certain variety of pine tree is 10.1 inches in three years. A biologist claims that a new variety will have a greater three-year growth. A random sample of 25 of the new variety has an average three-year growth of 10.8 inches and a standard deviation of 2.1 inches. The appropriate null and alternate hypotheses to test the biologist's claim are __________.

(a) H_0: $\mu = 10.8$ against H_a: $\mu > 10.8$
(b) H_0: $\mu = 10.8$ against H_a: $\mu \neq 10.8$
(c) H_0: $\mu = 10.1$ against H_a: $\mu > 10.1$
(d) H_0: $\mu = 10.1$ against H_a: $\mu \neq 10.1$

(1 mark)

Q12 Test of hypothesis H_0: $\mu = 50$ against H_1: $\mu > 50$ leads to __________.

(a) Left-tailed test
(b) Right-tailed test
(c) Two-tailed test
(d) Non of the mentioned

(1 mark)

Q13 The following are percentages of fat found in 5 samples of each of two brands of fruit yogurt:

A: 5.7, 4.5, 6.2, 6.3, 7.3
B: 6.3, 5.7, 5.9, 6.4, 5.1

Which of the following procedures is appropriate to test the hypothesis of equal average fat content in the two types of fruit yogurt?

(a) Paired t-test with 4 degrees of freedom
(b) Paired t-test with 5 degrees of freedom
(c) Two sample t-test with 8 degrees of freedom
(d) Two sample t-test with 9 degrees of freedom

(1 mark)

Q14 An advertising agency wants to test the hypothesis that the proportion of adults in Malaysia who read a Sunday Newspaper is 25 percent. The null hypothesis is that the proportion reading the Sunday Newspaper is __________.

(a) Equal to 25%
(b) Less than 25 %
(c) More than 25 %
(d) Different from 25%

(1 mark)

Q15 Given $\mu_0 = 130$, $\overline{X} = 150$, $\sigma = 25$ and $n = 4$; what test statistics is appropriate?

(a) F
(b) t
(c) χ^2
(d) Z

(1 mark)

Q16 A regression analysis is inappropriate when __________.

(a) there is heteroscedasticity in the scatter plot.
(b) the pattern of data points forms a reasonably straight line.
(c) you have two variables that are measured on an interval or ratio scale.
(d) you want to make predictions for one variable based on information about another variable.

(1 mark)

Q17 If the slope of the regression equation $Y = b_0 + b_1 X$ is positive, then __________.

(a) As x increases y decreases
(b) As x increases so does y
(c) As x decreases y remain
(d) As x decreases y increases

(1 mark)

Q18 If the regression equation is equal to $Y = 12.3 - 45.6X$, then 12.3 is the ________ while −45.6 is the ________ of the regression line.

(a) Intercept, slope
(b) Slope, intercept
(c) Radius, intercept
(d) Slope, regression coefficient

(1 mark)

Q19 If the correlation coefficient is a positive value, then the slope of the regression line ________.

(a) must also be positive
(b) can be either negative or positive
(c) can be zero
(d) cannot be zero

(1 mark)

Q20 Regression analysis was applied to the return rates of sparrowhawk colonies. Regression analysis was used to study the relationship between return rate (x: % of birds that return to the colony in a given year) and immigration rate (y: % of new adults that join the colony per year). The following regression equation was obtained.

$$y = 31.9 - 0.34x$$

Based on the above estimated regression equation, if the return rate were to decrease by 10% the rate of immigration to the colony would ________.

(a) increase by 34%
(b) increase by 3.4%
(c) decrease by 0.34%
(d) decrease by 3.4%

(1 mark)

PART B: SUBJECTIVE QUESTIONS

Q21 (a) A tire manufacturer reported the failure rate of tire quality and measured the tire lifetime. The test of selected tires mean lifetime is normally distributed with a mean of 60,000 miles and a standard deviation of 3500 miles. For a random sample of 30 tires,

(i) Identify the distribution's mean, standard deviation and skewness. Justify your answer.

(2 marks)

(ii) Find the probability that if you buy one such tire, it will last only 57,000 or fewer miles.

(4 marks)

(b) A cylindrical hole is drilled in a block in which the mean hole diameter is 9.70 mm and the variance is 0.35 mm. A cylindrical piston is placed in the hole where the mean diameter of the piston is 8.30 mm and the variance is 0.40 mm. Assume that both diameters are normally distributed.

(i) Find the probability that the piston will fit inside the hole with at least 0.10 mm clearance.

(6 marks)

(ii) Find the percentage of randomly selected piston and a block of hole will not fit together.

(5 marks)

(iii) If it is possible to adjust the mean hole diameter, determine the diameter value should be adjusted so that the clearance will be between 0.05 and 0.09 mm.

(3 marks)

Q22 (a) The engineer made 8 independent melting point measurements of silicon for IC design Assume that the obtained sample mean is 3410.14°C and sample standard deviation is 1.018.

(i) Find and interpret a 95% confidence for the mean melting point of the silicon.

(6 marks)

(ii) If the eight measurements were 3409.76, 3409.80, 3410.66, 3409.79, 3409.76, 3409.77, 3409.80 and 3409.78, would the confidence intervals in **Q22(a)(i)** be valid? Justify your answer.

(3 marks)

(b) The packaging time for a machine is known to be approximately normally distributed. A random sample of 8 packages is selected similarly from old and new machine. The sample from old machine gave an average packaging time of 42.14s with a sample standard deviation of 0.683. While the new machine gave an average packaging time of 43.23s with a sample standard deviation of 0.750.

(i) Construct a 95% confidence interval for the true variance of both types of machines.

(6 marks)

(ii) Based on the confidence interval as in **Q22(b)(i)**, evaluate is it reasonable to assume that the two populations sample have equal variances? Test the two-population variance at 0.05 level of significance and interpret your result.

(5 marks)

Q23 An experiment on the effect of fertilizer usage on jackfruit production has been conducted by a fertilizer company. The mean jackfruit weight harvested in a plantation area has a normal distribution with the mean weight is 23 kilograms with a standard deviation of 9 kilograms per fruit. The fertilizer company claims that their product can give heavier jackfruit production if using their new fertilizer product. A sample of 25 jackfruit was selected at random and the mean weight of the jackfruit weight is 19 kilograms per fruit. Analyze the fertilizer company's claim at 1% significant level.

(a) Identify the suitable null and alternative hypothesis, type of tests (right-tailed, left-tailed, or two-tailed) and distribution for this question.

(3 marks)

(b) Explain the reason to chose this distribution as solution to Q23(a).

(1 mark)

(c) Calculate the critical value and test statistic. Write what is the rejection region, decision rule, decision and its reason to accept or reject H_0.

(12 marks)

(d) Sketch the distribution graph for the hypothesis. Mark the critical value and test statistic. Label the "Reject the H_0" and "Do not reject H_0" areas.

(2 marks)

(e) Write the conclusion for the hypothesis.

(2 marks)

Q24 The number of hours spent per week watching TV, y, and the number of years of education, x, were recorded for 10 randomly selected individuals. The results are given in **Table Q24**.

Table Q24

Number of years of education, x	Number of hours spent per week watching TV, y
12	10
14	9
11	15
16	8
16	5
18	4
12	20
20	4
10	6
12	15

Given $\sum x = 141,\ \sum y = 106,\ \sum x^2 = 2085 \sum y = 1408,\ \sum xy = 1351$

(a) Find the least-squares line for these data.

(8 marks)

(b) Compute the sum square error for the data.

(4 marks)

(c) Compute the mean square error of data.

(3 marks)

(d) Compute the linear correlation coefficient.

(3 marks)

(e) Based on answer **Q24(d)**, interpret the value of linear correlation coefficient.

(2 marks)

– END OF QUESTIONS –

FINAL EXAMINATION

SEMESTER/SESSION : SEM. I / 2022/2023	PROGRAMME CODE : BEJ/ BEV
COURSE NAME : ENGINEERING STATISTICS / ENGINEERING MATHEMATICS V	COURSE CODE : BEE32502 / BEE31702

Formula

Random Variable:

$$\sum_{i=-\infty}^{\infty} P(x_i) = 1 \qquad E(X) = \sum_{\forall x} x \cdot P(x) \qquad E(X^2) = \sum_{\forall x} x^2 \cdot P(x)$$

$$\int_{-\infty}^{\infty} f(x)\,dx = 1 \qquad E(X) = \int_{-\infty}^{\infty} x \cdot P(x)\,dx \qquad E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot P(x)\,dx$$

$$Var(X) = E(X^2) - [E(X)]^2$$

Special Probability Distributions:

$$P(x=r) = {}^nC_r \cdot p^r \cdot q^{n-r},\ r = 0,\ 1,\ 2,\ \ldots,\ n \qquad X \sim B(n,p)$$

$$P(X=r) = \frac{e^{-\mu} \cdot \mu^r}{r!},\ r = 0,\ 1,\ 2,\ \ldots,\ \infty \qquad X \sim P_o(\mu)$$

$$Z = \frac{X-\mu}{\sigma} \qquad Z \sim N(0,\ 1) \qquad X \sim N(\mu,\ \sigma^2)$$

Sampling Distributions:

$$\overline{X} \sim N\left(\mu,\ \frac{\sigma^2}{n}\right) \qquad Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,\ 1) \qquad T = \frac{\overline{x}-\mu}{s/\sqrt{n}}$$

$$\overline{X}_1 - \overline{X}_2 \sim N\left(\mu_1 - \mu_2,\ \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

Estimations:

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2$$

$$(\overline{x}_1 - \overline{x}_2) - z_{\alpha/2}\left(\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) < \mu_1 - \mu_2 < (\overline{x}_1 - \overline{x}_2) + z_{\alpha/2}\left(\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

$$(\overline{x}_1 - \overline{x}_2) - z_{\alpha/2}\left(\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right) < \mu_1 - \mu_2 < (\overline{x}_1 - \overline{x}_2) + z_{\alpha/2}\left(\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right)$$

FINAL EXAMINATION

SEMESTER/SESSION : SEM. I / 2022/2023
COURSE NAME : ENGINEERING STATISTICS / ENGINEERING MATHEMATICS V
PROGRAMME CODE : BEJ/ BEV
COURSE CODE : BEE32502 / BEE31702

$$(\overline{x}_1 - \overline{x}_2) - t_{a/2,v} \cdot s_p^2 \left(\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right) < \mu_1 - \mu_2 < (\overline{x}_1 - \overline{x}_2) + t_{a/2,v} \cdot s_p^2 \left(\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right)$$

where Pooled estimate of variance is $S_P^2 = \dfrac{(n_1 - 1)s_1^2 + (n^2 - 1)s_1^2}{n_1 + n_2 - 2}$ with $v = n_1 + n_2 - 2$

$$(\overline{x}_1 - \overline{x}_2) - t_{a/2,v} \left(\sqrt{\frac{1}{n} (s_1^2 + s_2^2)} \right) < \mu_1 - \mu_2 < (\overline{x}_1 - \overline{x}_2) + t_{a/2,v} \left(\sqrt{\frac{1}{n} (s_1^2 + s_2^2)} \right)$$

with $v = 2(n-1)$

$$(\overline{x}_1 - \overline{x}_2) - t_{a/2,v} \left(\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right) < \mu_1 - \mu_2 < (\overline{x}_1 - \overline{x}_2) + t_{a/2,v} \left(\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right)$$

with $v = \dfrac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\frac{\left(\frac{s_1^2}{n_1} \right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2} \right)^2}{n_2 - 1}}$

$\dfrac{(n-1)s^2}{\chi^2_{a/2,\ v}} < \sigma^2 < \dfrac{(n-1)s^2}{\chi^2_{1-a/2,\ v}}$ with $v = n - 1$

$\dfrac{s_1^2}{s_2^2} \left(\dfrac{1}{F_{df_1, df_2, \frac{a}{2}}} \right) < \dfrac{\sigma_1^2}{\sigma_2^2} < \dfrac{s_1^2}{s_2^2} \left(F_{df_1, df_2, \frac{a}{2}} \right)$ with $v_1 = n_1 - 1,\ and\ v_2 = n_2 - 1$

Hypothesis Testing:

$$Z_{test} = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$T = \dfrac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ with $v = n_1 + n_2 - 2$

$$Z = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$T = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{n}(s_1^2 + s_2^2)}}$$

$T = \dfrac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ with $v = \dfrac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\frac{\left(\frac{s_1^2}{n_1} \right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2} \right)^2}{n_2 - 1}}$ $S_p^2 = \dfrac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2}$

$$\chi^2 = \frac{(n-1)\, s^2}{\sigma^2}$$

FINAL EXAMINATION

SEMESTER/SESSION : SEM. I / 2022/2023	PROGRAMME CODE : BEJ/ BEV
COURSE NAME : ENGINEERING STATISTICS / ENGINEERING MATHEMATICS V	COURSE CODE : BEE32502 / BEE31702

Simple Linear Regression:

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right) \qquad S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2$$

$$S_{yy} = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2 \qquad \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$\beta_1 = \frac{S_{xy}}{S_{xx}} \qquad \beta_0 = \overline{y} - \beta_1 \overline{x} \qquad y = \beta_0 + \beta_1 x$$

$$R^2 \frac{S_{yy} - SSE}{S_{yy}} = 1 - \frac{SSE}{S_{yy}} \qquad r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}}$$

$$SSE = S_{yy} - \beta_1 S_{xy} \qquad MSE = \frac{SSE}{n-2}$$

$$T = \frac{\beta_1 - \beta_c}{\sqrt{\frac{MSE}{S_{xx}}}} \sim t_{n-2} \qquad T = \frac{\beta_0 - \beta_c}{\sqrt{MSE\left(\frac{1}{n} + \frac{\overline{x}^2}{S_{xx}}\right)}} \sim t_{n-2}$$