AIR FORCE INSTITUTE OF TECHNOLOGY
FACULTY OF SCIENCE
STATISTICS
SECOND SEMESTER EXAMINATION 2020/2021
B.ENG. AEROSPACE/ AUTOMOTIVE/ MECHANICAL/ MECHATRONICS/ METALLURGICAL AND MATERIALS
100 & 200 LEVEL

Course Title:	INTRODUCTION TO STATISTICS
Course Code:	STA 102
Credit Unit:	2 Units
Instruction:	Answer question ONE and any other THREE questions.
Duration:	2 HOURS 30 MINUTES
Date:	3rd FEBRUARY, 2022

Question 1 (10 marks)

a. Define class interval **(1 mrk)**

b. List the three (3) types of class interval **($1\frac{1}{2}$ mrks)**

c. The following are mean stress in N/mm^2 of a passenger aircraft that cruises at 250m/s with an average block for length journey of 1500km for 40 days.

48	70	60	47	51	55	59	63
47	53	67	62	64	69	57	56
58	63	65	62	49	64	53	59
61	67	47	56	64	66	49	52
58	53	63	69	59	51	70	56

Construct a frequency table for the above data showing the following: *Class interval*, *Frequency*, *Class mark*, *Class boundary*, *Cumulative frequency*, and *Relative frequency*. **($7\frac{1}{2}$ mrks)**

Question 2 (20 marks)

a. Enumerate four (4) methods of data collection **(4 mrks)**

b. State any six (6) characteristics of a good questionnaire **(3 mrks)**

c. A Cu-Zn alloy has an initial grain diameter of 0.01mm. The alloy is then heated to various temperatures, permitting grain growth to occur. The distribution below shows the times required for the grains to grow to a diameter of 0.06mm; **(13 mrks)**

Temperature (°C)	90 – 100	101 – 111	112 – 122	123 – 133	134 – 144	145 – 155
Time (Minutes)	7	2	9	8	4	5

Determine the following:

i. Mean growth rate;
ii. Median and Mode; and
iii. Standard deviation

Question 3 (20 marks)

a. i. Distinguish clearly between primary and secondary data **(4 mrks)**

ii. State any two (2) merits and demerits each of the two sources of statistical data **(4 mrks)**

b. In quest for a new electrical component used for sound enhancement to gain market acceptance manufactured by XYZ Electric Limited. The company developed two (2) quality control systems (QCSs) to ensure each component that came-off a production line meets specification. The following results for each QCSs are given below: **(6 mrks)**

QCS	Average Number of Defective Components	Standard Deviation
A	8.6	15.2
B	8.17	4.44

i. Which QCS is more efficient in detecting defective components and give reason for your answer?
ii. Using the coefficient of variation, determine the ratio of the standard deviation to the mean for each quality control system.

c. A sample containing 400 items taken from the output of a production line. Defective items occurred randomly and independently at a rate of 0.016 of the entire items. What is the probability that the sample will have? **(6 mrks)**
   i. Less than 3 defective items
   ii. More than 2 defective items
   iii. What is the standard deviation of the number of defective items?

Question 4 (20 marks)

a. State any four (4) properties of the normal distribution **(4 mrks)**

b. Assume X is a normal random variable whose mean and variance are 12 and 5.2, respectively. Draw the normal curve and shade the following standardized values: **(5 mrks)**
   i. $F(Z \leq 3.6)$
   ii. $F(Z > 3.6)$
   iii. $F(-3.1 \leq Z \leq 2.5)$
   iv. $F(2.1 \leq Z \leq 3.0)$
   v. $F(-3.1 \leq Z \leq -1.5)$

c. Military radar and missile detection systems are designed to warn a country of an enemy attack. A reliability question deals with whether X detection system newly developed by a team of experts will be able to identify an attack and issue a warming. Assume that a particular detection system has a 0.90 probability of detecting a missile attack. Use the binomial probability distribution to answer the following questions. **(11 mrks)**
   i. What is the probability that a single detection system will detect an attack?
   ii. If two detection systems are installed in the same area and operate independently, the system will detect the attack.
   iii. If three detection systems are installed, what is the probability that at least one of the systems will detect the attack?
   iv. Would you recommend that single or multiple detection system(s) be used? Explain.

Question 5 (20 marks)

a. Define the following term: **(4 mrks)**
   i. Simple linear regression; and
   ii. Correlation analysis

b. The number of aircraft landing and the main tire consumption statistics of first quarter of each year from 2012 – 2021. Assuming that Y is the main tire consumption and X is the number of aircraft landing. **(16 mrks)**

Year	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Landing Frequency	623	504	745	346	672	239	656	543	294	592
Tire Consumption	70	54	83	48	78	26	54	76	39	62

   i. Establish a simple linear regression model and predict the main tire consumption if the landing frequency is 800 (Use the Ordinary Least Squares method).
   ii. What is the correlation coefficient between the main tire consumption and aircraft landing? (Use the Pearson's Correlation Coefficient).

Question 6 (20 marks)

a. i. Briefly explain the term "Data editing" **(2 mrks)**
   ii. List four (4) key factors to consider when editing primary data **(4 mrks)**

b. In a robotic contest, competitors were ranked by three judges in the following order

Judges	A	2	4	7	8	6	1	5	3
	B	8	2	6	1	5	3	7	4
	C	2	3	5	7	4	6	1	8

Use the Spearman's rank correlation coefficient to discuss which pair of judges has similar decisions in scoring the best participant with sound knowledge in designing robots. **(14 mrks)**

FORMULA SHEET FOR SECOND SEMESTER EXAM (2020/2021) ACADEMIC SESSION

MEASURES OF CENTRAL TENDENCY AND PARTITION

Mean $(\bar{X}) = \frac{\sum f_i x_i}{\sum f_i}$

Median $= L_m + \left(\frac{\frac{n}{2} - f_c}{f_c}\right) C$

Mode $= L_m + \left(\frac{f_m - f_b}{(f_m - f_b) + (f_m - f_a)}\right) C$

Quartile $(Q_i) = L_{q_i} + \left(\frac{\frac{in}{4} - cf_{q_i}}{f_{q_i}}\right) C$

Decile $(D_i) = L_{D_i} + \left(\frac{\frac{in}{10} - cf_{D_i}}{f_{D_i}}\right) C$

Percentile $(P_i) = L_{P_i} + \left(\frac{\frac{in}{100} - cf_{P_i}}{f_{P_i}}\right) C$

Measures of Dispersion

Variance $S^2 = \frac{\sum f(X - \bar{X})^2}{n-1}$

Standard Deviation $S = \sqrt{\frac{\sum f(X - \bar{X})^2}{n-1}}$

Quartile Deviation $Q.D = \frac{Q_3 - Q_1}{2}$

Coefficient of Variation $C.V = \frac{S}{\bar{X}} \times 100\%$

Correlation Analysis

Pearson's Correlation Coefficient:

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

Spearman's Rank Correlation:

$$\gamma_s = 1 - \frac{6\sum D^2}{n(n^2 - 1)}$$

$D = R_X - R_Y$

PROBABILITY DISTRIBUTIONS

Binomial Distribution:

$P(X = x) = \binom{n}{x} P^x (1 - P)^{n-x}$

$\mu = np, \quad \sigma = \sqrt{np(1-p)}$

$\binom{n}{x} = \frac{n!}{(n-x)!\,x!}$

Poisson Distribution:

$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$

$e = 2.71828$

$\mu = \lambda, \quad \sigma = \sqrt{\lambda}$

Normal Distribution:

$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

The standardized normal random variable

$Z = \frac{x - \mu}{\sigma}$

Ordinary Least-Squares Line (Linear Regression Line or Line of Best Fit)

$\hat{y} = \alpha + \beta x$

Note that, α is the intercept and β is the slope

Where $\alpha = \bar{y} - \beta\bar{x}$

$\bar{y} = \frac{\sum y}{n}$, $\bar{x} = \frac{\sum x}{n}$ and,

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