National University of Computer & Emerging Sciences, Karachi
Fall-2022 CS-Department
Final Examination, 2nd January 2023, 8:30 am – 11:30 am

Course Code: MT2005	**Course Name: Probability & Statistics**
Instructor Names: Mr. Osama Bin Ajaz and Mr. Muhammad Amjad	
Student Roll No:	**Section No:**

Instructions:

- Return the question paper.
- Read each question completely before answering it. There are **08 questions and 02 pages**
- In case of any ambiguity, you may make an assumption. But your assumption should not contradict any statement in the question paper.
- Show all steps of the solution. Otherwise marks will be deduct accordingly.
- All the answers must be solved according to the sequence given in the question paper.
- Answer written using pencil will not have checked. Pencil is only allowed to draw diagrams or write program code.

Time: **3** hrs. **Total Points** = **70** (50% weightage)

Question # 01: [CLO-1] Max Marks: 06

A computer assembly Company receives 30% of parts from supplier X, 30% of parts from suppliers Y, and the remaining 40% of the parts from supplier Z. One percent of parts supplied by X, 3 percent of parts supplied by Y, and 2 percent of parts supplied by Z are defective. If an assembled computer has a defective part in it, what is the probability that this part was received from Supplier Z? (Hint: apply Bayes' rule)

Question # 02: [CLO-2] Max Marks: 08

A program consists of two modules. The number of errors, X, in the first module and the number of errors Y, in the second module have the following Joint probability distribution in the table. Find:

f(x , y)		Y		
		0	1	2
X	0	0	2/70	3/70
	1	3/70	18/70	9/70
	2	9/70	18/70	3/70
	3	3/70	2/70	0

(i) The marginal distribution of X and Y. **[2]**

(ii)The probability of no errors in the first Module **[1]**

(iii) The distribution of errors in the program $P[\,X + Y \le 2]$ **[2]**

(iv)$P(X = 0/Y = 2)$,**[2]**

(v) Also, find out if errors in the two modules X and Y occurs independently? **[1]**

Question # 03 [CLO-3] Max Marks: 08

The following data show the numbers of defects in 100,000 lines of code in a particular type of software program developed in two different software houses. Is there enough evidence to claim that there is a significant difference between the programs developed in the two countries? Assume population variances to be equal (Critical value t = 2.04)

Software house 01:	48	39	42	52	40	48	52	52
	54	48	52	55	43	46	48	52
Software house 02:	50	48	42	40	43	48	50	46
	38	38	36	40	40	48	48	45

Question # 04 [CLO-3] Max Marks: 08

Several neurosurgeons wanted to determine whether a dynamic system (Z-plate) reduced the operative time relative to a static system (ALPS plate). The data displayed in the table below on operative times, in minutes, for the two systems. At the 5% significance level, does the data provide sufficient evidence to conclude that the mean operative time is less with the dynamic system than with the static system? Assume population variances are unequal.
[critical value t = -1.74]

Dynamic: 370, 360, 510, 445, 295, 315, 490, 345, 450, 505, 335, 380, 325, and 500.
Static: 430, 445, 455, 455, 490, and 535.

Question # 05: [CLO-3] Max Marks: 06

A study was conducted to determine if the performance of a certain type of surgery on young horses had any effect on certain kinds of blood cell types in the animal. Fluid samples were taken from each of six foals before and after surgery. The samples were analyzed for the number of postoperative white blood cell (WBC) leukocytes. A preoperative measure of WBC leukocytes was also measured. The data are given as follows:

Foal	**1**	**2**	**3**	**4**	**5**	**6**
Pre- surgery	10.80	12.90	9.59	8.81	12.00	6.07
Post-surgery	10.60	16.60	17.20	14.00	10.60	8.60

(all values are x 10^{-3})

Use paired test to determine if there is a significant change in WBC leukocytes with the surgery. *(critical value 2.57)*

Question # 06: [CLO-3] Max Marks:12

A manufacturer suspects that the batches of raw material furnished by his supplier differ significantly in calcium content. There are a large number of batches currently in the warehouse. Five of these are randomly selected for study. A chemist makes five determinations on each batch and obtains the following data:

Batch 1	Batch 2	Batch 3	Batch 4	Batch 5
23.46	23.59	23.51	23.28	23.29
23.48	23.46	23.64	23.40	23.46
23.56	23.42	23.46	23.37	23.37
23.39	23.49	23.52	23.46	23.32
23.40	23.50	23.49	23.39	23.38

Is there significant variation in calcium content from batch to batch. Apply analysis of variance technique and use 5% level of significance. (Critical value 2.86).

Question # 07: [CLO-3] Max Maks:10

The Statistics Consulting Center analyzed data on normal woodchucks for the Department of Veterinary Medicine. The variables of interest were body weight in grams and heart weight in grams. It was desired to develop a **linear regression** equation in order to determine if there is a significant linear relationship between heart weight and total body weight. Also construct **scatter plot** and calculate **correlation**.

Body Weight	4050	2465	3120	5700	2595	3640	2050	4235	2935	4975	36900	2800	2775	2170	2370	2055	2025	2645	2675
Heart Weight	11.2	12.4	10.5	13.2	9.8	11	10.8	10.4	12.2	11.2	10.8	14.2	12.2	10	12.3	12.5	11.8	16	13.8

Question # 08: [CLO-3] Max marks: 12

A client from the Department of Mechanical Engineering approached the Consulting Center at Virginia Tech for help in analyzing an experiment dealing with gas turbine engines. The voltage output (y) of engines was measured at various combinations of blade speed and sensor extension.

Volts	1.95	2.5	2.93	1.69	1.23	3.13	1.55	1.94	2.18	2.7	1.32	1.6	1.89	2.15	1.09	1.26	1.57	1.92
Speed	6336	7099	8026	6230	5369	8343	6522	7310	7974	8501	6646	7384	8000	8545	6755	7362	7934	8554
Extension	0	0	0	0	0	0	0.006	0.006	0.006	0.006	0.012	0.012	0.012	0.012	0.018	0.018	0.018	0.018

Fit a **multiple linear regression model**.

Useful Information:

ANOVA : $SS_T = \sum\sum y_{ij}^2 - \frac{y_{..}^2}{N}$

$$\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2 = n\sum_{i=1}^{k}(\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2$$

$$SSTR = n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + n_3(\bar{x}_3 - \bar{x})^2 + n_4(\bar{x}_4 - \bar{x})^2$$

$$SS_{\text{Treatments}} = \sum_{i=1}^{a} \frac{y_{i.}^2}{n_i} - \frac{y_{..}^2}{N}$$

Multiple regression Model:

Y = α + $\beta_1 X_1$ + $\beta_2 X_2$ + ε

Normal Equation:

$$\sum Y = na + b_1\sum X_1 + b_2\sum X_2$$
$$\sum X_1 Y = a\sum X_1 + b_1\sum X_1^2 + b_2\sum X_1 X_2$$
$$\sum X_2 Y = a\sum X_2 + b_1\sum X_1 X_2 + b_2\sum X_2^2$$

Simultaneous solution to above normal equations give values of b_1 & b_2 as:

$$b_1 = \frac{(\Sigma X_1 Y)\left(\Sigma X_2^2\right) - (\Sigma X_2 Y)(\Sigma X_1 X_2)}{\left(\Sigma X_1^2\right)\left(\Sigma X_2^2\right) - (\Sigma X_1 X_2)^2}$$

$$b_2 = \frac{(\Sigma X_2 Y)\left(\Sigma X_1^2\right) - (\Sigma X_1 Y)(\Sigma X_1 X_2)}{\left(\Sigma X_1^2\right)\left(\Sigma X_2^2\right) - (\Sigma X_1 X_2)^2}$$

t-test: (a) $\frac{\bar{x}-\mu}{s/\sqrt{n}}$ **, t =** $\frac{\bar{X}_1 - \bar{X}_2 - \mu_1 - \mu_2}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ **t =** $\frac{\bar{X}_1 - \bar{X}_2 - \mu_1 - \mu_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$

Coefficient of correlation= r = $\frac{n\sum YX - \sum Y\sum X}{\sqrt{n\sum X^2 - (\sum X)^2}\sqrt{n\sum Y^2 - (\sum Y)^2}}$

Multiple correlation coefficient: **R =** $\sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1}.r_{yx_2}.r_{x_1x_2}}{1 - r_{x_1x_2}^2}}$

Standard Error = $\sqrt{\frac{\sum(y-y')^2}{n-2}}$

$\boldsymbol{Sample\ variance} = s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$

$\boldsymbol{Population\ variance} = \sigma^2 = \frac{\Sigma(x-\bar{x})^2}{n}$

hint Q2iv: use: $f\left(\frac{x}{y}\right) = P(X = x/_{Y = y})\}$

Bayes' Rule: $P(B_r|A) = \frac{P(B_r \cap A)}{\sum_{i=1}^{k} P(B_i \cap A)} = \frac{P(B_r)P(A|B_r)}{\sum_{i=1}^{k} P(B_i)P(A|B_i)},$