National University of Computer and Emerging Sciences

Lahore Campus

Probability & Statistics (MT2005)

MID-1 Exam

Date: Mar 01, 2025

Total Time (Hrs): 1

Total Marks: 35

Total Questions: 2

ALL SECTIONS **BCS,BSE**

Course Instructor(s)

Ms. Sarah Ahmad

Ms. Kanwal Saleem

Ms. Huma Akbar

Dr. Nuzhat Aftab

_____ _____ _____

Roll No Section Student Signature

Do not write below this line

Strictly Attempt all questions on Question paper. You may use rough sheet but DO NOT attach it.

Lead Pencil work wouldn't be marked or claimed for rechecking. Use permanent ink pen.

If you found any ambiguity in the data then do not ask anything to the invigilator, just make an assumption and continue solving your paper.

Q2 requires a direct answer. Avoid details, cutting, or overwriting, as it will be marked zero.

CLO 1: **Statistical Data Interpretation**: Analyze and interpret various data types by computing measures of central tendency and dispersion, constructing frequency distributions, and utilizing graphical techniques for precise data representation.

Q1: **[Marks:10+15]**

Case Study: Computer System Performance Metrics

This dataset provides performance insights for 15 different computer systems, including hardware specifications and real-time usage statistics for in-depth evaluation.

Variables in the Dataset:

- **System ID**: Unique identifier for each computer system.
- **System Type**: Whether the system is Personal, Workstation, or Server**.**
- **Storage Type**: The type of storage used **(**HDD, SSD, and Hybrid**).**
- **Total Storage (GB)**: Total available storage capacity.
- **Power Consumption (Watts)**: Power used by the system during normal operation.
- **CPU Utilization (%)**: Percentage of CPU usage during operation.
- **GPU Available**: Whether the system has a dedicated GPU **(**Yes/No**).**
- **Network Speed (Mbps)**: The maximum internet speed the system can handle.
- **Disk Utilization (%)**: Percentage of used disk space.
- **RAM (GB)**: Amount of installed memory.

Sample Data of Computer System Performance (15 Records)

System ID	System Type	Storage Type	Total Storage	Power	CPU Utilization	GPU Available	Network Speed	Disk	RAM
1	Personal	SSD	512	150	35	Yes	1000	60	16
2	Workstation	HDD	1024	250	70	Yes	500	75	32
3	Server	Hybrid	2048	400	85	Yes	1000	80	64
4	Personal	SSD	256	100	40	No	300	50	8
5	Workstation	HDD	512	180	60	Yes	800	65	16
6	Server	Hybrid	4096	500	90	Yes	2000	85	128
7	Personal	SSD	1024	200	50	Yes	2205	70	32
8	Workstation	HDD	2048	300	75	Yes	900	78	64
9	Server	Hybrid	8192	600	95	Yes	2500	90	256
10	Personal	SSD	512	130	55	No	1900	58	16
11	Workstation	HDD	1024	220	72	Yes	600	76	32
12	Server	Hybrid	4096	480	88	Yes	1500	84	128
13	Personal	SSD	256	90	30	No	200	40	8
14	Workstation	HDD	512	190	65	Yes	700	72	16
15	Server	Hybrid	8192	700	98	Yes	3000	92	256

Tasks:

a) To effectively classify network performance, group 'Network Speed' into classes starting at 180 with a class width of 600. Draw cross-tabulation to compare these 'Network Speed' classifications with 'System Type' for meaningful insights.

b) Detect anomalies in 'Power Consumption' by computing z-scores. Additionally, show that the z-scores of Power Consumption follows a standard normal distribution with a mean of 0 and a standard deviation of 1.

Q1 Task (a) *Solve here*

Network Speed	System Type		
	Personal	Workstation	Server
180-779	2	3	0
780-1379	1	2	1
1380-1979	1	0	1
1980-2579	1	0	2
2580-3179	0	0	1

System Type	Network Speed				
	180-779	780-1379	1380-1979	1980-2579	2580-3179
Personal	2	1	1	1	0
Workstation	3	2	0	0	0
Server	0	1	1	2	1

Each cell 0.5 as inside box all entries 7.5 and 0.5 each correct class interval 2.5 (Total =10)

CLO 1: **Statistical Data Interpretation**: Analyze and interpret various data types by computing measures of central tendency and dispersion, constructing frequency distributions, and utilizing graphical techniques for precise data representation.

Q2: **[Marks:10]**

Sr.	SCENARIOS	ANSWER	10
1	A researcher records the daily count of security breaches in a cloud server system for six months to analyze trends in cyber-attacks. What type of study design is this?	Time Series	**1**
2	A sample of 50 website visitors is selected from a total user base of 10,000, but the selection missed users from a specific region. What type of error is this?	Sampling Error	**1**
3	IP addresses assigned to devices are an example of which scale of measurement?	Nominal	**1**
4	The battery percentage displayed on smartphones represents which scale of measurement?	Ratio	**1**
5	A network provider wants to classify internet speeds for customer plans. The raw data consists of speeds in Mbps (e.g., 10 Mbps, 50 Mbps, 100 Mbps).What is the nature of the variable if internet speed is categorized into groups such as Slow (0–10 Mbps), Moderate (11–50 Mbps), Fast (51–100 Mbps), and Ultra-Fast (101+ Mbps)?	Categorical/Qualitative	**1**
6	Consider the response times (milliseconds) of a web server, where lower and upper quartile are 32 ms and 45 ms respectively with a median of 35 ms. Identify the shape of the distribution of response time.	Positively Skewed	**1**
7	A cloud computing company monitors server response times (in milliseconds) for two different setups; Server A: Mean = 200 ms, Standard Deviation = 20 ms and Server B: Mean = 500 ms, Standard Deviation = 50 ms. Calculate the relative variation in response times for both servers and comment which server has more consistent response times.	○ Server A ○ Server B ✓ **Both are Same**	**1**
8	A computer science department analyzes exam scores from 100 students in two different courses; Course A: Scores are mostly high, with a few students scoring very low (Mean = 85, Median = 95). Course B: Scores are more evenly spread, with some students scoring both very high and very low (Mean = 70, Median = 70). The department wants to assess the distribution of scores to understand if the performance in each course is symmetric or skewed. Comment Accordingly.	Course A: Negatively Skewed	**0.5**
		Course B: Symmetric	**0.5**
9	A software testing team categorizes detected errors into: Syntax Errors:50, Logic Errors:30, Runtime Errors:20. Which type of graph is most suitable for visualizing these categories?	Bar Chart/Pie Chart	**1**
10	A data center tracks CPU usage for 20 servers and provides the following five-number summary: Min = 20, Q_1= 35, Q_2= 45, Q_3= 55, Max = 95. Determine the threshold value used to flag anomalies and identify on which side the anomalies lie (above or below).	Threshold: 85 upper fence	**0.5**
		Side: above	**0.5**

Q1 Task (b) Detect anomalies in 'Power Consumption' by computing z-scores. Also, comment on it with reason. Additionally, show that the z-scores of Power Consumption follows a standard normal distribution with a mean of 0 and a standard deviation of 1.

Solve here:

Measures	Marks = 10
Sum X = 4490	1
Sum X^2 = 1857300 or Sum $(X - 299.33)^2$ = 513293.3	2
$Mean_X$ = 299.33	1
SDx = 191.48	1
Sum Z = 0.0002	1
Sum Z^2 =13.99942	2
Mean Z = 0.000013 ~ 0	1
SDz = 0.99998 ~ 1	1

Measures	Marks = 10
$Mean_X$ = 299.33	2
SDx = 191.48	3
Mean Z = 0.000013 ~ 0	2
SDz = 0.99998 ~ 1	3
Marks including working. 1 for each answer and remaining working	

Sort in Ascending order	
X_i	Z_i (4-decimal)
90	-1.0932
100	-1.0410
130	-0.8843
150	-0.7799
180	-0.6232
190	-0.5710
200	-0.5187
220	-0.4143
250	-0.2576
300	0.0035
400	0.5257
480	0.9435
500	1.0480
600	1.5702
700	2.0925
0.25 each z-score (Total 3.75) **0.25 for z formula.**	

- ✓ **All the z scores lie b/w -3 to +3.**
- ✓ **Hence No outlier detected.**

1

Working:

Mean of X = $\frac{\sum X}{n} = \frac{4490}{15} = 299.3333$

SD of X = $\sqrt{\frac{1}{n-1}\left[\sum X_i^2 - \frac{(\sum X_i)^2}{n}\right]}$

$= \sqrt{\frac{1}{14}\left[1857300 - \frac{(4490)^2}{15}\right]}$

= 191.47796

Mean of Z = $\frac{\sum Z}{n} = \frac{0.0002}{15} = 0.000013 \sim 0$

SD of Z = $\sqrt{\frac{1}{n-1}\left[\sum Z_i^2 - \frac{(\sum Z_i)^2}{n}\right]}$

$= \sqrt{\frac{1}{14}\left[13.99942 - \frac{(0.0002)^2}{15}\right]}$

$= 0.99998 \sim 1$

Note: If wrong variables are picked up by the students in a and b parts then allocate 3 marks in a part and 5 marks in b part for correct working.