

Tribhuvan University  
Institute of Science and Technology  
**SCHOOL OF MATHEMATICAL SCIENCES**  
Second Assessment 2081

*Subject: Research Methodology*  
*Course No: MDS 601*  
*Level: MDS / II Year / III Semester*

*Full marks: 45*  
*Pass Marks: 22.5*  
*Time: 2hrs*

*Candidates are required to give their answers in their own words as far as practicable.*  
**Attempt All Questions.**

**Group A [5x3=15]**

1. What is bias how can it be reduced while interviewing with respondent?
2. Explain how stratified sample can reduce the standard error?
3. Describe the nature and purpose of case study design?
4. Define literature review. What are the benefits of good literature survey?
5. Explain the ordinal and ratio scale with suitable examples.

**Group B [5x6=30]**

6. A research scholar has to work as a judge and derive the truth not as a Pleader who is only eager to prove his/her case in favor of his/her plaintiff. Discuss the statement pointing out the objective of research.
7. A manager notices the number of grievances in the organization is increasing. The manager wishes to investigate this occurrence. What research design seems appropriate for the study? Give appropriate explanation.
8. Under what condition might a field experiment be better than a laboratory experiment?
9. Define the term reliability. Discuss the various types of reliability.

**OR**

Discuss the relative advantages and disadvantages of questionnaire and schedule.

10. What are the various methods of documenting sources? Describe the APA method of citation with examples.

**OR**

What is a research proposal? Point out the different steps in research proposal with brief explanation.

\*\*\*

Tribhuvan University  
Institute of Science and Technology  
**SCHOOL OF MATHEMATICAL SCIENCES**  
Second Assessment 2081

*Subject: Advanced Data Mining*  
*Course No: MDS 602*  
*Level: MDS / II Year / III Semester*

*Full marks: 45*  
*Pass Marks: 22.5*  
*Time: 2hrs*

*Candidates are required to give their answers in their own words as far as practicable.*  
**Attempt All Questions.**

**Group A [5x3=15]**

1. What is warehousing? What are the sources of data for building a data warehouse?
2. What limitations of OLTP are solved in OLAP so that it can be used for data mining?
3. List out the different techniques that can be applied for data mining with use cases.
4. What is KDD? List the possible tasks that can be done in the data pre-processing phase.
5. Write different steps of the K-NN algorithm.

**Group B [5x6=30]**

6. What is Clustering? Explain giving DB Scan as an example algorithm.

**OR**

7. What is the Apriori Algorithm? Discuss with its use case.
7. Explain data visualization techniques.

**OR**

Explain the working of SVM with examples.

8. What is Ensemble techniques? Explain with example.
9. What is an outlier? How can clustering be used for outlier detection?
10. How can we build the Association based recommendation system? Explain with examples.

\*\*\*

Tribhuvan University  
Institute of Science and Technology  
**SCHOOL OF MATHEMATICAL SCIENCES**  
Second Assessment 2081

**Subject: Techniques for Big Data**  
**Course No: MDS 603**  
**Level: MDS / II Year / III Semester**

**Full marks: 45**  
**Pass Marks: 22.5**  
**Time: 2hrs**

*Candidates are required to give their answers in their own words as far as practicable.*  
**Attempt All Questions.**

**Group A [5x3=15]**

1. What is NoSQL? How is it different than SQL?
2. How Apache Spark is different from Hadoop Mapreduce?
3. Explain different execution modes of Apache Pig.
4. Explain about HBase architecture briefly.
5. What is the process to create UDFs (User Defined Functions) in Apache Pig?

**Group B [5x6=30]**

6. Explain about the CAP theorem. What is mean by Eventual 6. 6. Consistency in NoSQL databases.

**OR**

What are the different types of NoSQL databases? Explain with examples.

7. Write short notes on:
  - a) Spark SQL
  - b) Spark MLlib
8. Explain architecture of Apache Spark. Also explain about spark transformations and actions with examples.
9. Explain about Hive Shell, Hive Services and Hive Metastore.

**OR**

Explain about Hive architecture in detail. Also, compare Apache Hive with Apache Pig.

10. Write mongodb query for the following database:

➤ db.order.find({})

```
[
  { "customer": "Alice", "product": "Laptop", "price": 1200, "date": "2024-03-01" },
  { "customer": "Bob", "product": "Phone", "price": 800, "date": "2024-03-01" },
  { "customer": "Alice", "product": "Tablet", "price": 500, "date": "2024-03-02" },
  { "customer": "Charlie", "product": "Laptop", "price": 1000, "date": "2024-03-02" },
  { "customer": "Bob", "product": "Laptop", "price": 1300, "date": "2024-03-03" },
  { "customer": "Alice", "product": "Phone", "price": 700, "date": "2024-03-03" },
  { "customer": "Bob", "product": "Tablet", "price": 600, "date": "2024-03-03" }
]
```

- a) Write a query to find maximum price per customer.
- b) Write a query to find total revenue per day.
- c) Write a query to find average price per customer per product.

\*\*\*



Tribhuvan University  
Institute of Science and Technology  
**SCHOOL OF MATHEMATICAL SCIENCES**  
Second Assessment 2081

**Subject: Decision Analysis**  
**Course No: MDS 606**  
**Level: MDS / II Year / III Semester**

**Full marks: 45**  
**Pass Marks: 22.5**  
**Time: 2hrs**

*Candidates are required to give their answers in their own words as far as practicable.*  
**Attempt All Questions.**

**Group A [5x3=15]**

1. Distinguish between Risk Appetite and Risk tolerance.
2. Determine the saddle point solution, the associated pure strategies and the value of the game for the following game. The payoff matrix for player A is given by

Player A	Player B			
	B <sub>1</sub>	B <sub>2</sub>	B <sub>3</sub>	B <sub>4</sub>
A <sub>1</sub>	8	6	2	8
A <sub>2</sub>	8	9	4	5
A <sub>3</sub>	7	5	3	5

3. An office equipment manufactures produces two kinds of products chair and lamp. Production of either requires one hour of production capacity in the plant. The plant has a maximum production capacity of 50 hours per week because of the limited sales capacity, the maximum number of chairs and lamps that can be sold are 6 and 8 respectively. The gross margin from sale of chair is 90 and 60 for a lamp. The plant manager desires to determine the number of units of each product that should be produced per week in consideration of the following equally ranked goals.  
**Goal 1:** Available production capacity should be utilized as much as possible but not exceeded.  
**Goal 2:** Sales of two products should be as much as possible.  
**Goal 3:** Overtime should not exceed 20% of available production time.  
 Formulate the problem as a Goal programming.
4. A man has the choice of running either a hot -snack stall or an ice-cream stall at a sea side resort during the summer season. If it is a fairly cool summer, he should make Rs.5000 by running the hot-snack stall, but if the summer is quite hot, he can only expect to make Rs.1000. On the other hand, if he operates the ice-creams stall, his profit is estimated at Rs.6500 if the summer is hot, but only Rs.1000 if it is cool. There is a 40% chance of the summer being hot. What should be his decision to maximize expected profit by using EMV criterion?
5. Explain about decision making Conditions.

**Group B [5x6= 30]**

6. Find the optimal strategies using the dominance rule for player A and B in the following game. Also obtain the value of the game.

Player A's Strategy	Player B's Strategy		
	B <sub>1</sub>	B <sub>2</sub>	B <sub>3</sub>
A <sub>1</sub>	6	12	7
A <sub>2</sub>	11	7	12
A <sub>3</sub>	10	6	11

7. A company manufactures two products radio and transistors which must be processed through assembly and finishing departments. Assembly has 90 hours available; finishing can handle up to 72 hours of work. Manufacturing one radio requires 6 hours in assembly and 3 hours in finishing. Each transistor requires 2 hours in assembly and 4 hours in finishing. If profit is 1.20 per radio and 1.90 per transistor, determine the best combination of radios and transistors to realize a maximum profit of 2000. Formulate a problem as a GP problem and solve it.

OR

A company manufactures two products radio and transistors which must be processed through assembly and finishing departments. Assembly has 90 hours available; finishing can handle up to 72 hours of work. Manufacturing one radio requires 6 hours in assembly and 3 hours in finishing. Each transistor requires 2 hours in assembly and 4 hours in finishing. If profit is 1.20 per radio and 1.90 per transistor, determine the best combination of radios and transistors to realize a maximum profit of 2000. Formulate a problem as a GP problem and solve it.

8. What is Group decision making? Describe the techniques of group decision making.  
*OR what is Decision Analysis? Describe different types of decision theories with suitable Examples?*
9. A newspaper vendor buys a new started local paper at the rate of Rs. 5 and sells it at the rate of Rs. 10. The unsold papers do not have any value. The vendor knows that he cannot sell more than 20 papers in a day and the minimum sale would not be less than 18. How many papers should he buy based on  
 a) maximax criterion b) maximin criterion and c) minimax regret criterion?
10. A beer distributor buys kegs for Rs.8 each and sells them for 12 each. All the kegs left at the end of the day are worthless. Following is the distribution of sales during 100 days observations.

Kegs sold	20	21	22	23	24	Total
No. of Days	5	20	30	35	10	100

- a) Find the optimal quantity that can maximize the expected profit.  
 b) Find the expected value of perfect information.  
 c) What is the cost of uncertainty?

OR

A physician purchases a particular vaccine on Monday each week. The vaccine must be used within the week following, otherwise it becomes worthless. The vaccine costs Rs.2 per dose and the physician charges Rs.4 per dose. In the past 50 weeks, the physician has administered the vaccine in the following quantities.

Doses per week	20	25	50	60
Number of weeks	5	15	25	5

- a) Determine how many doses the physician should buy every week.  
 b) Compute EVPI.  
 c) A physician is thinking of spending on a small market survey to obtain additional information regarding the demand levels. How much should he be willing to spend on such a survey?

\*\*\*



Tribhuvan University  
Institute of Science and Technology  
**SCHOOL OF MATHEMATICAL SCIENCES**  
Second Assessment 2081

*Subject: Monte Carlo Methods*  
*Course No: MDS 607*  
*Level: MDS / II Year / III Semester*

*Full marks: 45*  
*Pass Marks: 22.5*  
*Time: 2hrs*

*Candidates are required to give their answers in their own words as far as practicable.*

**Attempt All Questions.**

**Group A [5x3=15]**

1. What is Bayesian inference? Explain.
2. Describe motivating vignettes of Statistical inferences.
3. Illustrate meaning of "Reparametrization".
4. Distinguish prior and posterior distributions of Bayesian Statistics.
5. Explain briefly the meaning of "Gibbs Sampling".

**Group B [5x6=30]**

6. Prove Metropolis-Hastings Algorithm to carry on any MCMC.

**OR**

Give an account of changes carried out by Hastings in Metropolis algorithm.

7. Discuss convergence criteria in Gibbs sampling.

**OR**

Consider bivariate normal distribution. Discuss necessary theory and Algorithm to solve this problem using Gibbs sampling.

8. Explain algorithm to apply Gibbs Sampling to study Hierarchical model.
9. How do we check for convergence of a Metropolis-Hastings sampler? Discuss with an example.
10. A manufacturer claims that the shipment contains only 5% of defective items, but the inspector feels that in fact it is 10%. We have to decide whether to accept or to reject the shipment based on  $\theta$ , the proportion of defective parts. Before we see the real data, let's assign a 50-50 chance to both suggested values of  $\theta$  i.e.  $\pi(0.05) = \pi(0.10) = 0.5$ . A random sample of 20 parts has 3 defective ones. Calculate the posterior distribution of  $\theta$ .

\*\*\*