Tribhuvan University
Institute of Sciences and Technology

SCHOOL OF MATHEMATICAL SCIENCES

First Assessment 2080

Subject: **Programming with Python**
Course No: **MDS 551**
Level: **MDS /I Year /II Semester**

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

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

Group A [5 ×3=15]

1. Why do we need variables in programming? What is global variable?
2. Explain membership operators with example.
3. Write a program to test whether a number entered is leap year or not.
4. Write a program to count number of digits in a string.
5. Write a program using function to find sum of first n natural numbers.

Group B [5 × 6 = 30]

6. Explain for loop and while loop with example.

   OR

   Write a program to display prime numbers up to 100.
7. Explain list data type with example. How list is different from tuple? What is list comprehension?

   OR

   Explain dictionary data type with example. What is nested dictionary?
8. List some benefits of using functions. Explain different ways of passing arguments in functions.
9. How do you read and write files using Python program. Explain the process of reading and writing CSV files in Python with suitable program.
10. Explain the use of break and continue statements in programming with example. What is Pass statement?

***

Tribhuvan University

Institute of Sciences and Technology

SCHOOL OF MATHEMATICAL SCIENCES

First Assessment 2080

Subject: **Applied Machine Learning**

Course No: **MDS 552**

Level: **MDS /I Year / II 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 [5×3=15]

1. How supervised learning is a costly approach compared to unsupervised?
2. Differentiate between batch and stochastic gradient descent.
3. Describe precision and recall as evaluation metrics for classification task.
4. Write an algorithm for K-Means Clustering.
5. Explain Soft Max Regression over Logistic Regression with suitable example.

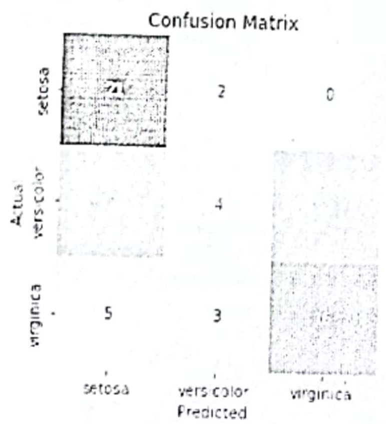
Group B [5×6=30]

6. Derive the form of Logistic Regression for classification tasks.
7. Explain the various form of evaluation metrics used for regression.
8. What do you understand by ensemble learning? Explain its types.

OR

Consider the following data points $(1,\ 0), (0,\ 1), (-1,\ 0), (0,\ -1)$ as negatively labelled data and (8, 10), (10, 8), (12, 10), (10, 12) as positively labelled data. Determine the equation of hyper plane that divides the above data points into two classes using SVM and predict in which class (8, 6) belongs.

9. Given a confusion matrix from Iris flower classification as below, compute f1 score and interpret the result.



10. Execute hierarchical clustering in the provided data as below where rows and columns contains the point and values are the distances between the point. Also, draw dendrogram to illustrate diagrammatically.

Points	1	2	3	4	5
1	0	9	3	6	11
2	9	0	7	5	10
3	3	7	0	9	2
4	6	5	9	0	5
5	11	10	2	5	0

OR

Briefly describe how DBSCAN algorithm works. Apply DBSCAN algorithm to the above given data.

***

Tribhuvan University

Institute of Science and Technology

SCHOOL OF MATHEMATICAL SCIENCE

First Assessment 2080

Subject: **Statistical Method for Data Science** *Full Marks*: 45

Course No: **MDS 553** *Pass Mark*: 22.5

Level: **MDS / I year/ II Semester** *Time*: 2hrs.

Attempt all the questions

Group A [5×3=15]

1. Define generalized power series distribution and show that the negative binomial distribution is a special case of GPSD.

2. The number of virus infected computers of five different capacity of hard disk is given below:

Capacity of hard disk(GB)	500	320	1000	2000	400
No. of virus infected	11	15	20	3	1

Test whether the computers of five hard disk are uniformly infected using Kolmogorov Smirnov test.

3. An important part of the customer service responsibilities of a telephone company relates to the speed with which troubles in residential service can be repaired. Suppose past data indicate that the likelihood is 0.70 that troubles in residential service can be repaired on the same day. For the first five troubles reported on a given day, what is the probability that
   a) All five will be repaired on the same day?
   b) At least three will be repaired on the same day?
   c) Fewer than two will be repaired on the same day?

4. What do you mean by multinomial distribution and find the moment generating function of the distribution.

5. The average number of claims per hour made to the Gnecco and Trust Insurance Company for damages or losses incurred in moving is 3.1. What is the probability that in any given hour
   a) Fewer than three claims will be made?
   b) Exactly three claims will be made?
   c) Three or more claims will be made?

Group B [5×6=30]

6. Define negative binomial distribution and find the mean and variance of the distribution.

OR

Define Binomial distribution and find the mean and variance of the distribution.

7. Define errors in hypothesis testing. Ten accountants were given intensive coaching and four tests were conducted in a month. The scores of tests 1 and 4 are given below.

Accountants	1	2	3	4	5	6	7	8	9	10
Marks in 1^{st} test	50	42	51	42	60	41	70	55	62	38
Marks in 4^{th} test	62	40	61	52	68	51	64	63	72	50

Does the score from 1^{st} to 4^{th} test shows an improvement using Wilcoxon Matched pairs signed rank test? Test at the 5% level of significance.

8. An IQ test was given to a random sample of 15 male and 20 female students of a university. Their scores were recorded as follows:

Male: 56, 66, 62, 81, 75, 73, 83, 68, 48, 70, 60, 77, 86, 44, 72

Female: 63, 77, 65, 71, 74, 60, 76, 61, 67, 72, 64, 65, 55, 89, 45, 53, 68, 73, 50, 81

Use median test at 0.01 level of significance to determine whether IQ of male and female students are same in the university.

9. State and prove Neymann-Pearson's Lemma.

OR

For testing $H_0:\theta=1.5$ against $H_1:\theta=2.5$ on the basis of single observation drawn from uniform distribution with probability density function,

$$f(x) = 1/\theta;\ 0<X<\theta$$

Calculate the probabilities of type first error, type second error and power of the test, if critical regions are (i) $w=\{x:0.8 \leq x\}$ and (ii) $w=\{x:1.5 \leq x \leq 2\}$.

10. The following data represents the operating times in hours for three types of scientific pocket calculators before a recharge is required:

Calculator A	4.9	6.1	4.3	4.6	5.3		
Calculator B	5.5	5.4	6.2	5.8	5.5	5.2	4.8
Calculator C	6.4	6.8	5.6	6.5	6.3	6.6	

Use Kruskal-Wallis H test, at the 0.05 level of significance, to test the hypothesis that the operating times for all three calculators are equal.

***

Tribhuvan University
Institute of Sciences and Technology
SCHOOL OF MATHEMATICAL SCIENCES

First Assessment 2080

Subject: **Multivariable Calculus for Data Science**
Course No: MDS 554
Level: MDS /I Year /II Semester

Full Marks: 45
Pass Marks: 22.5
Time: 2.00 hrs

Candidates are required to give their answer in their own words as far as practicable. All questions carry equal marks.

Group A [5×3=15]

1. State geometrical meaning of scalar triple product. Find the volume of the parallelepiped determined by the vectors $\vec{a} = \vec{i} + 2\vec{j} + 3\vec{k}$, $\vec{b} = -\vec{i} + \vec{j} + 2\vec{k}$, and $\vec{c} = 2\vec{i} + \vec{j} + 4\vec{k}$ [0.5+2.5]

2. Find the parametric and symmetric equation of the straight line through (2, 1,0) and perpendicular to both the vectors $\vec{i} + \vec{j}$ and $\vec{j} + \vec{k}$. [1.5+1.5]

3. Define normal and bi-normal vectors of a vector function of scalar variable t.
Find the bi-normal vector of the space curve $\vec{r} = x\vec{i} + y\vec{j} + z\vec{k}$, where $x = t^2, y = t^2, z = t^3$ at point (1, 1, 0). [1.5+1.5]

4. Find the domain and range of the function $f(x,y) = \sqrt{4 - x^2 - y^2}$. Sketching a Contour map, describe the level curves of the function for the values c = 0, 1, 2, 3.
Find the limit, if it exists, or show that the limit does not exist at indicated point:
(a) $f(x, y) = \dfrac{x^2 y}{x^4 + y^2}$ at (0, 0)
(b) $f(x, y) = \dfrac{xy^3}{x + y}$ at (−1, 2) [1.5+1.5]

5. Show that the function $u = e^x \sin y + e^y \cos x$ satisfies Laplace's equation
$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0.$$ [3]

Group B [5×6=30]

6. Find the vector equation of a straight line through the two vectors $\vec{a}$ and $\vec{b}$. Find an equation of the plane (by vector method) through the points whose position vectors are $3\vec{i} - \vec{j} + 2\vec{k}$, $8\vec{i} + 2\vec{j} + 4\vec{k}$ and $-\vec{i} - 2\vec{j} - 3\vec{k}$ [3+3]

7. Derive the formula for the derivative of vector triple product $\frac{d}{dt}[\vec{r_1} \times (\vec{r_2} \times \vec{r_3})]$ of three vectors $\vec{r_1}, \vec{r_2}$ and $\vec{r_3}$. If $\vec{r_1} = a\cos t\,\vec{i} + b\sin t\,\vec{j}$, $\vec{r_2} = -a\sin t\,\vec{i} + b\cos t\,\vec{j} + t\vec{k}$ and $\vec{r_3} = \vec{i} + 2\vec{j} + 3\vec{k}$, find $\frac{d}{dt}[\vec{r_1} \cdot (\vec{r_2} \times \vec{r_3})]$. [3+3]

OR

Define curvature of the vector function of scalar variable $\vec{r} = \vec{r}(t)$. Find the curvature of the curves

(a) $x = a(\theta + \sin\theta)$, $y = a(1 - \cos\theta)$ at point $\theta = 0$. [3]

(b) $y = a \log \sec \dfrac{x}{a}$ at any point (x, y). [3]

8. Find the domain and range of the function $f(x, y) = \sqrt{4-x^2-y^2}$. Describe the graph of f. Sketch a contour map of this surface using level curves corresponding to $c = 0, 1, 2, 3, 4$. **[1.5 + 1.5 + 1.5 + 1.5]**

9. **(a)** Let $f(x, y)$ be defined on an open disk D that contains the point (a, b). Prove that if the functions f_{xy} and f_{yx} are continuous on D, then $f_{xy}(a, b) = f_{yx}(a, b)$. **[4]**

(b) Let $f(x, y) = e^x \cos y$. Confirm that the mixed second-order partial derivatives of f are the same. **[2]**

10. **(a)** Prove that if a function $z = f(x, y)$ is differentiable at a point(a, b), then it is continuous at the point. **[3]**

(b) Show that $f(x, y) = xe^{xy}$ is differentiable at $(1, 0)$ and find its linearization there. Then use it to approximate $f(1.1, -0.1)$. **[3]**

OR

(a) Prove that if f is a differentiable function of x and y, then f has a directional derivative at (x_0, y_0) in the direction of any unit vector $u = (a, b)$ and

$$D_u f(x_0, y_0) = f_x(x_0, y_0)a + f_y(x_0, y_0)b.$$ **[3]**

(b) Find the closest points from the origin to the curve $x^2y = 16$. **[3]**

□□

Tribhuvan University
Institute of Science and Technology
SCHOOL OF MATHEMATICAL SCIENCES
First Assessment 2080

Subject: **Natural Language Processing**
Course No.: **MDS 555**
Level: **MDS / I Year/ II Semester**

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

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

Attempt ALL questions.

Group A [5 × 3 = 15]

1. What is NLP? List out the challenges of NLP?
2. What are the differences between Stammer and Lemmatizer ?
3. Define HMM in context of NLP.
4. What is Parsing in NLP? Discuss with examples.
5. Discuss the usability of the Word Net.

Group B [5 × 6 = 30]

6. What is Context Free Grammar (CFG)? "Probabilistic CFG solve the ambiguity". Justify this statement with examples.

OR

Write a note on POS tagging.

7. Define Finite State Machines (FSM). How can we use FSM to analyze presence of the prefix? Explain with examples.

OR

What is n-gram language model? Explain the use of the n-gram model in NLP task.

8. How can we use NLP techniques to solve the Name Entity Recognition(NER) related problems. Explain.
9. What is Semantic Analysis? Explain, the elements of Semantic Analysis with example.
10. What is the pipeline(steps) that we follow while developing NLP based solutions?

***

Tribhuvan University

SCHOOL OF MATHEMATICAL SCIENCES

First Reassessment Exam 2080

Subject :Natural Language Processing
Level: **MDS/IYear/IISemester**
Course No : **MDS505**

Full Mark:**45**
Pass Mark:**22.5**
Time:**2Hours**

Candidates are required to give their answer in their own words as far as practicable. The figures in the margin indicate full marks.

Attempt ALL question.

Group A [5×3=15]

1. Explain the applications of NLP.
2. What is Stemmer? Discuss which problem of stemmer is solved by Lemmatizer ?
3. Define HMM in context of NLP.
4. Explain dependency parsing in NLP.
5. What is Word Net? What type of data can be found on Word Net?

Group B [5×6=30]

6. Explain Probabilistic CFG with its application.

   OR

   Write a note on Name Entity Recognition(NER).
7. Define Finite State Machines (FSM). How can we use FSM to analyze presence of the prefix? Explain with examples.

   OR

   Explain Word Sense Disambiguation(WSD)with its applications.
8. Explain the steps involved in NLP based problem solving.
9. List the elements of Semantic Analysis. What is Semantic Analysis? What are the difference between Polysemy and Homonymy? Explain with examples.
10. Explain the use of then-gram model in NLP task.

***

Tribhuvan University
Institute of Sciences and Technology
SCHOOL OF MATHEMATICAL SCIENCES

First Re-Assessment 2080

Subject: **Multivariable Calculus for Data Science**
Course No: MDS 554
Level: MDS /**I Year /II Semester**

Full Marks: 45
Pass Marks: 22.5
Time: 2.00hrs

Candidates are required to give their answer in their own words as far as practicable. All questions carry equal marks.

Group A [5×3]

1. Determine whether the points $A(1, 3, 2)$, $B(3, -1, 6)$, $C(5, 2, 0)$, and $D(3, 6, -4)$ lie in the same plane. **[3]**
2. If $\vec{u} + \vec{v}$ and $\vec{u} - \vec{v}$ are orthogonal, show that the vectors $\vec{u}$ and $\vec{v}$ must have the same length. Also write a vector equation for the line through the point $(1, 0, 6)$ and perpendicular to the plane $x + 3y + z = 5$. **[2+1]**
3. Define curvature of a vector function $\vec{r} = \vec{r}(t)$. Find the curvature of the vector function $\vec{r}(t) = a\cos t\,\vec{i} + a\sin t\,\vec{j}$, where a is constant. **[1+2]**
4. Find the domain and range of the function $f(x, y) = \dfrac{y}{\sqrt{x}}$. **[3]**
5. Show that the function $u = \ln(x + ct)$ satisfies the wave equation
$$\frac{\partial^2 u}{\partial t^2} = c^2\frac{\partial^2 u}{\partial x^2}.$$ **[3]**

Group B [5 ×6]

6. (a) Find a vector equation of a straight line through the given vector $\vec{a}$ and parallel to the vector $\vec{b}$. **[3]**
   (b) Find a vector equation and parametric equation for the line through the point $(1, 0, 6)$ and perpendicular to the plane $x + 3y + z = 5$. **[3]**
7. (a) Write the formula for the derivative of scalar triple product of three vectors $\vec{r}_1, \vec{r}_2$ and $\vec{r}_3$. If $\vec{r}_1 = a\cos t\,\vec{i} + b\sin t\,\vec{j}$, and $\vec{r}_2 = -a\sin t\,\vec{i} + b\cos t\,\vec{j} + t\vec{k}$, find $\dfrac{d}{dt}(\vec{r}_2 \times \vec{r}_1)$. **[1+2]**
   (b) Evaluate the derivative of $\dfrac{\vec{r}}{r}$ w.r.t t. **[1]**
   (c) If $\dfrac{d\vec{a}}{dt} = \vec{c} \times \vec{a}$ and $\dfrac{d\vec{b}}{dt} = \vec{c} \times \vec{b}$, show that: $\dfrac{d}{dt}(\vec{a} \times \vec{b}) = \vec{c} \times (\vec{a} \times \vec{b})$. **[2]**

OR

7. (a) Find the tangent, normal and bi-normal vector of the space curve $\vec{r} = x\vec{i} + y\vec{j} + z\vec{k}$ where $x = t, y = t^2, z = t^2$ at point $(1, 1, 1)$. **[3]**
   (b) Find the radius of curvature at any point ϕ for the parametric curve $x = a\cos\phi, y = b\sin\phi$. **[3]**

8. Find $\lim_{(x,\ y)\to(0,0)} \frac{3x^2y}{x^2+y^2}$ if it exists. [6]

9. (a) Prove that if $x = x(t)$, $y = y(t)$ and $z = f(x, y)$ are differentiable functions, then $z = f(x(t), y(t))$ is a differentiable function of t and

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t},$$

where the ordinary derivatives are evaluated at t and the partial derivatives are evaluated at (x, y). [4]

(b) If $z = f(x, y)$ and f has continuous second order partial derivatives and $x = r^2 + s^2$, $y = 2rs$, find $\frac{\partial z}{\partial r}$, $\frac{\partial^2 z}{\partial r^2}$. [2]

10. Find and classify all the critical points of $f(x, y) = 4 + x^3 + y^3 - 3xy$. [3]

OR

(a) Assume that $f(x, y)$ and $g(x, y)$are differentiable functions. Prove that if $f(x, y)$ has a local minimum or a local maximum on the constraint curve $g(x, y) = 0$ at $P = (a, b)$, and if $\nabla g_P \neq 0$, then there is a scalar λ such that $\nabla f_P(x, y) = \lambda \nabla g_P(x, y)$. [3]

(b) Find the point on the sphere $x^2 + y^2 + z^2 = 1$ farthest from (1, 2, 3). [3]

□□

100/2080
2024/4/9

Tribhuvan University
Institute of Sciences and Technology
SCHOOL OF MATHEMATICAL SCIENCES
First Assessment 2080 (Re-exam)

Subject: **Programming with Python** | *Full Marks*: **45**

Course No: **MDS 551** | *Pass Marks:* **22.5**

Level: **MDS /I Year /II Semester** | *Time*: **2 hrs**

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

Group A [5 ×3=15]

1. Define variable. Explain global variable with example.
2. Explain identity operators with example.
3. Write a program to test whether a number entered is leap year or not.
4. Write a program to count number of whitespaces in a string.
5. Write a program using recursive function to find factorial of a number.

Group B [5 × 6 = 30]

6. Explain different forms of if statements with example.

   OR

   Write a program count even and odd numbers stored in a list.
7. Explain list data type with example. How list is different from tuple? What is list comprehension?

   OR

   Explain dictionary data type with example. What is nested dictionary?
8. List some benefits of using functions. What is return statement? Explain lambda function with example.
9. Explain the process of reading and writing CSV files with suitable program.
10. Why do we need looping in programming? What is while loop? What are the uses of break statement?

***

for x in r

Tribhuvan University

Institute of Sciences and Technology

SCHOOL OF MATHEMATICAL SCIENCES

First Reassessment 2080

Subject: **Applied Machine Learning**

Course No: **MDS 552**

Level: **MDS /I Year / II 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 [5×3=15]

1. Explain in brief the concept of Reinforcement Learning.
2. Describe how gradient descent is used as optimization algorithm.
3. What do you mean by cross validation? How is it useful in obtaining better model during machine learning?
4. Explain the concept of Gaussian Mixture algorithm.
5. Describe the random forest as machine learning algorithms.

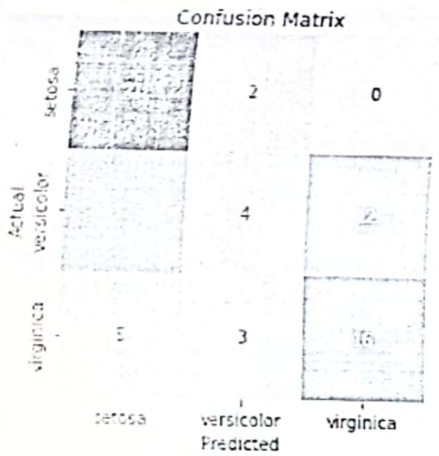
Group A [5×6=30]

6. Derive the form of weight update rule for linear regression.
7. Explain the concept of Support Vector Machine. Describe different kernels used in Support Vector Machine with their mathematical form?

OR

Explain over fitting and under fitting in machine learning with suitable example. How can we mitigate them.

8. Explain how bagging and boosting are useful in obtaining efficient machine learning model.
9. Given a confusion matrix from Iris flower classification as below, compute precision and recall for each class and interpret the result.



10. Execute hierarchical clustering in the provided data as below where rows and columns contains the point and values are the distances between the point. Also, draw dendrogram to illustrate diagrammatically.

Points	1	2	3	4	5
1	0	9	3	6	11
2	9	0	7	5	10
3	3	7	0	9	2
4	6	5	9	0	5
5	11	10	2	5	0

OR

Briefly describe how DBSCAN algorithm works. Apply DBSCAN algorithm to the above given data.

***

Tribhuvan University
Institute of Science and Technology
SCHOOL OF MATHEMATICAL SCIENCE
first Reassessment 2080

Subject: **Statistical Method for Data Science**
Course No: **MDS 553**
Level: **MDS / I year/ II Semester**

Full Marks: 45
Pass Mark: 22.5
Time: 2hrs

Candidates are required to give their answer in their own words as far as practicable.
Attempt all the questions

Group A [5×3=15]

1. Define generalized power series distribution and show that the binomial distribution is a special case of GPSD.
2. Differentiate between parametric and non-parametric test.
3. A multiple choice test has 5 questions. There are 4 choices for each question. A student, who has not studied for the test, decides all the answer of all questions randomly. What is the probability that he will get
   a) At most two questions correct? b) At least one questions correct?
4. Obtain the correlation coefficient of multinomial distribution.
5. Nepal Rastra bank is responsible for printing the country's paper money. It has an impressively small printing error only 0.5 percent of all bills are too flawed for circulation. What is the probability that out of a batch of 1000 bills,
   a) None are too flawed for circulation. b) Ten are too flawed for circulation.
   c) Fifteen are too flawed for circulation.

Group B [5×6=30]

6. Define binomial distribution and find the mean and variance of the distribution.

OR

Define Poisson distribution and find the mean and variance of the distribution.

7. Define null and alternative hypothesis.
   The same C programming papers were marked by two teachers A and B. The final score were recorded as follows:

Teacher A	73	89	82	43	80	73	66	45	50	55
Teacher B	88	78	91	48	85	74	77	31	40	28

   Using median test at 5% level of significance to determine if the marks distribution of two teachers differ significantly.
8. A random sample of size 25 is drawn from a normal population with mean μ and standard deviation 3. In testing $H_0: \mu=20$ against $H_1: \mu>20$ it is decided that H_0 will be rejected if the sample mean is greater than 21.4. Calculate (i) probability of type I error (ii) probability of type II error when (a) μ=21 and (b) 22.
9. In a certain computer hardware manufacturing industry six different types of machines are working to cut pieces of wires. The number of wires of unequal length recorded in a day is as follows:

Machine	1	2	3	4	5	6
No. of wire	2	0	4	8	5	11

   Do these data provide sufficient evidence that the six machines equally cut the wires of unequal length? Apply Kolmogorov Smirnov test at 5% level of significance.

OR

The heart beating rate of 5 vegetarians and 5 non vegetarians are recorded below:

Vegetarians	56	67	82	60	75
Non vegetarians	53	42	75	58	65

Is the mean heart beating rate of non-vegetarians significantly high? Use Mann Whitney U test.

10. A survey was conducted in four hospitals in a Kathmandu to obtain the number of babies born over a 12 months period. This time period was divided into four seasons to test the hypothesis that the birth rate is constant over all the four seasons. The results of the survey were as follows:

Hospital	**No. of births**			
	Winter	**Spring**	**Summer**	**Fall**
A	92	72	94	77
B	15	16	10	17
C	58	71	51	62
D	19	26	20	18

Analyze the data using Friedman two way ANOVA test.

***