Name: Enrolment No:	UPES UNIVERSITY WITH A PURPOSE

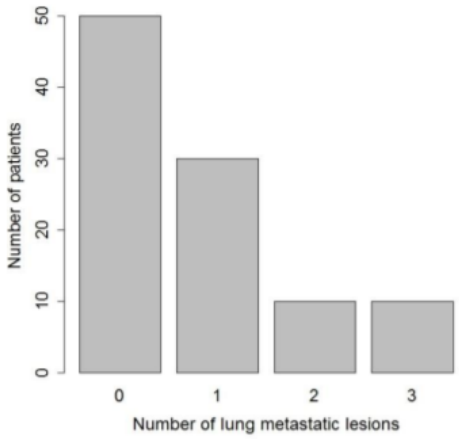
UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
Online End Semester Examination, June 2021

Course: Data Management Technologies — **Semester: II**
Program: M.Sc. (Clinical Research) — **Time: 03 hrs.**
Course Code: CSEG 7007P — **Max. Marks: 100**

SECTION-A

1. **Each question carries 1.5 Marks.**
2. **Select the correct answer / Complete the statement(s) / Fill in the blank(s).**

S. No.		CO
Q1	"You measure whether patients with diabetes who tell you that they are trying to lose weight can control their blood sugar better than patients who are not trying." This is an example of: (A) Observational study (B) Interventional study (C) Both (D) Neither Observational nor Interventional study	**CO1**
Q2	When designing a registry or a clinical research database a good practice is to: (A) Collect what you can gather first, think of the hypotheses later. (B) Think about the data quality process before you develop a hypothesis. (C) Let your hypotheses shape your data collection strategy. (D) None of the mentioned, the value of a good research database is not related to any specific scientific hypothesis.	**CO1**
Q3	The primary goal of Case Report Forms (CRF) design is to collect all the data required by the protocol in such a way that it can be analyzed according to the protocol and statistical analysis plan. (A) True (B) False	**CO1**
Q4	Read the following statements and choose the correct option: 1. Adverse event data (AE data) is just like other clinical data. 2. AE data is collected through case report forms (CRFs) or electronic CRFs (eCRFs) and stored along with the other subject data in the database. 3. Regular coding of reported terms is an important task of data management. 4. There is no relationship between serious AE reports in the data management system against those in the company's serious AE system. (A) 1, 2, and 3 only are True (B) 1, 3, and 4 only are True (C) 2, 3, and 4 only are True (D) 1, 2, 3, and 4 are True	**CO1**

Q5	Adverse event information for clinical trials can be grouped into three collection types or categories. What are those categories? 1. ---------- 2. ---------- 3. ----------	**CO1**
Q6	A Standard Operating Procedure (SOP) defines which of the following? 1. What the task is? 2. Who carries it out? 3. When it is to be carried out? 4. How it is to be carried out? 5. What evidence shows that it was carried out? (A) 1, 2, and 3 (B) 1, 3, and 5 (C) 2, 3, and 4 (D) 1, 2, 3, 4 and 5	**CO1**
Q7	In an experimental study, patients with advanced breast cancer are treated with a new drug. After three years of follow-up, radiographic scans are used to determine the number of metastatic lesions for each study patient. The number of lung metastases for each subject is shown in the figure. Based on these data, what is the average number of metastatic lung lesions for patients treated with the new drug?  (A) Between 0 and 1 (B) 1 (C) Between 1 and 2 (D) 2	**CO2**
Q8	Sample size depends on (A) Type of problem investigated (B) Resources available (C) Required precision (D) All of the above	**CO2**

Q9	Continuous variables are represented by (A) Histogram (B) Line Diagram (C) Bar Chart (D) Pie Chart	**CO2**
Q10	A …… observation is a time measure on a subject who does not have the outcome/event under study.	**CO2**
Q11	Normal distribution is symmetric about …… (A) mean (B) variance (C) standard deviation (D) covariance	**CO2**

Q12	We have submitted the following PROC SORT step, which generates an output dataset.
	`proc sort data = AV.employee out = employee;`
	`  by Designation;`
	`run;`
	In which library is the output dataset stored?
	(A) WORK
	(B) SASHELP
	(C) SASUSER
	(D) AV
		CO3

Q13	Employee
	name | age
	Ankit | 25
	Faizan | 24
	Kunal | 31
	Sunil | 30
	Salary
	name | salary
	Kunal | 60000
	Kunal | 55000
	Sunil | 37000
	Sunil | 34000
	As shown two tables above. How many variables would be in table "AV" after executing the below SAS program?
	`data AV;`
	`  merge Employee Salary;`
	`  by name;`
	`  totsal + salary;`
	`run;`
	(A) 2
	(B) 3
	(C) 4
	(D) 5
		CO3

Q14	Consider the following SAS Program:
	`data finance.earnings;`
	`Amount=1000;`
	`Rate=.075/12;`
	`do month=1 to 12;`
	`Earned+(amount+earned)*(rate);`
	`end;`
	`run;`
	What would be the value of month at the end of data step execution?
		CO3

Q15	If a variable in a SAS code contains letters or special characters, it can be numeric data type. (A) True (B) False	**CO3**

Q16	What happens if you submit the following program?
	`porc print data=work.newsalesemps;`
	`run;`
	(A) SAS does not execute the step.
	(B) SAS assumes that the keyword PROC is misspelled and executes the PROC PRINT step.
		CO3

Q17	Which of these projects would be best suited for Next Generation Sequencing? (A) To determine if a tumor sample contains a common missense mutation (B) To find the transcriptome of a tumor sample (C) To genotype ten genomic DNA samples for a known single nucleotide polymorphism (D) All of the above.	**CO4**
Q18	…… is the platform to download the genome sequences.	**CO4**
Q19	Which of the following is NOT true about Ab Initio–based programs for Gene Prediction? (A) The goal of the ab initio gene prediction programs is to discriminate exons from noncoding sequences. (B) The goal is joining exons together in the correct order. (C) The main difficulty is correct identification of exons. (D) To predict exons, the algorithms rely solely on gene signals.	**CO4**
Q20	A **Blank 1** is the entire sequence of nucleotides in the DNA that is in all of the chromosomes of a cell. Whereas a **Blank 2** is a small piece of the **Blank 1.** Where **Blank 1** = ……... **Blank 2** = ……...	**CO4**

SECTION-B 1. **Each question carries 5 Marks.** 2. **Instruction: Write short / brief notes in approx. 200 words.**		
Q1	Discuss the characteristics of Standard Operating Procedure (SOP) for clinical trials.	**CO1**
Q2	Write a short note on hypothesis testing.	**CO2**
Q3	Write a short note on gene prediction.	**CO4**
Q4	Differentiate between Do While and Do Until loop in SAS programming with suitable example.	**CO3**
SECTION-C **Instruction: Each question carries 15 Marks. Distribution of marks is also given.**		
Q1	Write SAS code to create two tables using PROC SQL by the name Student_B1 and Student_B2. Student_B1 will have variables as SAP Id, Stu_name, Course, and Branch. Student_B2 will have variables as SAP Id, Stu_name, Course, and Age. **[6]** Insert 5 observations in each of the table. **[3]** Further, join these two tables using left outer join, right outer join, and inner join based on Course variable. **[6]**	**CO3**
Q2	Explain Kaplan-Meier (K-M) estimate for survival analysis. **[5]** Find the K-M estimator and draw curve for the following data (n = 10): 2, 2, 3+, 5, 5+, 7, 9, 16, 16, 18+, where '+' means censored value. **[10]**	**CO2**
SECTION-D 1. **Each question carries 10 Marks. Distribution of marks is also given.** 2. **Instruction: Write long answers in approx. 300 words.**		
Q1	Describe the data exchange and data sharing workflow process in the context of clinical trials.	**CO1**
Q2	Describe the features of Next Generation Sequencing. **[2]** Discuss the genome sequence with its formats. **[2]** Mention the steps involved to search and download a specific genome sequence. **[6]**	**CO4**