2nd Tutorial Examination-2022
Department of Statistics
Course Code: STAT-401, Course Title: Statistical Inference II

Time: 1:00 Hour **Marks: 20**

1. Imagine a factory where the time taken by a machine to produce a widget is believed to be exponentially distributed with a parameter δ. After observing the machine's performance, it's deduced that the time t to produce a widget is governed by:

$$f(t \mid \delta) = 3\delta e^{-3\delta t}, \qquad t>0$$

From previous data, the factory manager believes that the parameter δ itself has an exponential distribution with a parameter of 4. Given a recorded production time t, what is the posterior distribution of δ?

2. What do you mean by conjugate prior? Let $X_1, X_2, \ldots, X_n$ be a random sample from Pareto density, find out the conjugate prior.

3. Define Jaffrey's' Non-Informative prior. Let $X \sim U(0,\theta)$, find Jeffrey's prior for θ.

4. Describe LINEX loss function. What is the relationship among LINEX, modified LINEX and entropy loss function.

CamScanner

Department of Statistics and Data Science
Jahangirnagar University
3rd Tutorial Examination-2022
Course Name:Statistical Inference II Course Code:STAT-401

Total Marks: 20 **Time: 1 hour**

a) Define Credible interval. Let $x_1, x_2, \cdots, x_n$ be a random sample from $N(0,\sigma^2)$. Find a large sample confidence interval for σ with an approximate confidence coefficient $1-\alpha$.

b) Suppose $X \sim N(\theta, 3)$. Assume that H_0, H_1 are equally likely. To test $H_0: \theta = 0 \; v.s. \; H_1: \theta = 1$,

i. find Bayes factor in favor of H_0 and interpret the result.

ii. If $n = 15, \bar{x} = 1.5$, obtain Posterior odds in favor of H_1.

c) Let X have the pmf $f(x;\theta) = \theta^x (1-\theta)^{1-x}$, $x = 0,1$, zero elsewhere. To test the hypothesis $H_0 : \theta = \frac{1}{4}$ vs. $H_1 : \theta < \frac{1}{4}$ by taking a random sample of size 10, show that, the best critical region is $\sum_{i=1}^{10} x_i \le 1$. Also find the power function $\gamma(\theta)$, $0 < \theta \le \frac{1}{4}$ and draw the power curve.

Course Title: Multivariate Analysis

Full marks: 20

a) Let $X_1, X_2, \ldots, X_{15}$ be 15 random sample from $N(\theta, 25)$. Find pitman location estimator of θ.

b) Let $x_1, x_2, \cdots, x_n$ be a random sample from $N\left(2, \sigma^2\right)$. Find a large sample confidence interval for σ^2 with an approximate confidence coefficient $1-\alpha$.

CS CamScanner

Department of Statistics
Jahangirnagar University
Part IV B.Sc. (Hons.) 1st Tutorial Examination 2022
Course Title: Multivariate Analysis, Course No. STAT-402

Time: 60 minutes | Full marks: 20

1. What is a multivariate normal distribution? How can you assess the assumption of multivariate normality? Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are iid $N_p(\boldsymbol{\mu}, \Sigma)$. Find the maximum likelihood estimate (MLE) of $\boldsymbol{\mu}$ and Σ. 5

2. If $\mathbf{X}_1, \mathbf{X}_2$ and $\mathbf{X}_3$ are jointly normal with quadratic form 10

$$Q = 7x_1^2 + 8x_2^2 + 6x_3^2 + 3x_1x_2 - 9x_1x_3 + 2x_2x_3 + x_1 - 7x_2 + 2x_3.$$

(i) Find the mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$

(ii) Find the conditional distribution of $\mathbf{X}_1$ given $2\mathbf{X}_1 - \mathbf{X}_2 + 3\mathbf{X}_3$.

(iii) Are $\mathbf{X}_3$ and $\mathbf{X}_1 + \mathbf{X}_2$ independently distributed? Explain.

3. What are the most common multivariate quality control charts? Define them with their uses. Construct a T^2-chart for data in Table 1. Use $\alpha = 0.01$. Hence, comment. 5

Table 1: Two measurements of stiffness with bending strength.

x_1	1232	1115	2205	1897
x_2	4175	6652	7612	10914

Where x_1 = stiffness and x_2 = bending strength are two measurements in pounds/(inches)^2 for a sample of 4 pieces of a particular grade of lumber.

4. A wildlife ecologist measured x_1 = tail length (in millimeters) and x_2 = wing length (in millimeters) for a sample of $n = 5$ female Hook-billed kites (a bird in the family Accipitridae). These data displays in the following.

x_1 (tail length)	191	197	180	180	196
x_2 (wing length)	284	285	273	276	288

Using the above data,

(i) Evaluate T^2 for testing $H_0 : \boldsymbol{\mu}' = [190 \quad 275]$.

(ii) Hence, find out the sampling distribution of T^2.

-Good Luck-

Department of Statistics
Jahangirnagar University
Part IV B.Sc. (Hons.) 2nd Tutorial Examination 2022
Course Title: Multivariate Analysis
Course No. STAT-402

Time: 90 minutes **Full marks: 20**

1. **(a)** Let the standardized variables $\mathbf{Z}' = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_p]$ have $\text{Cov}(\mathbf{Z}) = \boldsymbol{\rho}$ with the eigenvalue-eigenvector pairs $(\lambda_1, e_1), (\lambda_2, e_2), \ldots, (\lambda_p, e_p)$ where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. Hence, the i[th] principal component is $Y_i = \mathbf{e}_i'\mathbf{Z} = \mathbf{e}_i'(\mathbf{V}^{1/2})^{-1}(\mathbf{X} - \boldsymbol{\mu})$, $i = 1, \ldots, p$, where $\mathbf{V}^{1/2}$ is the diagonal standard deviation matrix. Then show that

   (i) $\sum_{i=1}^{p} Var(Y_i) = \sum_{i=1}^{p} Var(\mathbf{Z}_i) = p$

   (ii) $\rho_{Y_i, Z_k} = e_{ik}\sqrt{\lambda_i}$, $\quad i, k = 1, \ldots, p$.

   (b) In a study of size and shape relationships for painted turtles, Jolicoeur and Mosimann (1960) measured carapace length, width, and height. They performed a principal component analysis using logarithms of the dimensions of 24 male turtles. Following are the results of PCA

```
Importance of components:
                            PC1     PC2       PC3
Standard deviation     0.002262 0.00042 4.683e-19
Proportion of Variance 0.966680 0.03332 0.000e+00
Cumulative Proportion  0.966680 1.00000 1.000e+00

              PC1         PC2        PC3
Length -0.7616419 -0.08114189 -0.6428979
Width  -0.4550873 -0.63930114  0.6198303
Height -0.4612996  0.76466336  0.4499919
```

   Find and interpret the correlation coefficient between the 1st principal component and the Height of turtles.

2. **(a)** Define an orthogonal factor model with its components and assumptions. How could you check the adequacy of an orthogonal factor model?

   (b) The following R output is obtained for conducting the factor analysis with 5 variables and $m = 2$ common factors

```
Call:
factanal(x = x, factors = 2, method = "mle", scale = T, center = T)

Uniquenesses:
   V1    V2    V3    V4    V5
0.497 0.252 0.474 0.610 0.176

Loadings:                              Factor1 Factor2
   Factor1 Factor2        SS loadings       1.671   1.321
V1 0.601   0.378          Proportion Var    0.334   0.264
V2 0.849   0.165          Cumulative Var    0.334   0.598
V3 0.643   0.336
V4 0.365   0.507
V5 0.207   0.884

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 0.58 on 1 degree of freedom.
The p-value is 0.448
```

   Find the following:

   (i) Find the matrix of specific variances. Hence, define the most significant variable which fits neatly into this factors model.

   (ii) Find the estimated factor loadings and communalities. What proportion of the total population variance is explained by the first common factors?

-Good Luck-

Department of Statistics, Jahangirnagar University
Part IV B.Sc. (Hons.) 3rd Tutorial Examination 2022
Course Title: Multivariate Analysis, Course No. STAT-402

Time: 50 minutes **Full marks: 20**

1. Derive the allocation rule based on the minimum expected cost of the misclassification rule to separate objects into two multivariate normal populations π_1 and π_2 with mean vectors μ_1 and μ_2, and covariance matrices Σ_1 and Σ_2, respectively. **10**

2. Three psychological measurements x_1, x_2, and x_3 were taken on subjects in 2 neurotic groups (π_1 = Anxiety, and π_2 = Psychopathy). The sample means are given in Table 1.1. The pooled covariance matrix is given by

$$S_p = \begin{bmatrix} 2.1 & 0.2 & 0.3 \\ 0.2 & 1.6 & 0.1 \\ 0.3 & 0.1 & 2.5 \end{bmatrix}$$

Table 1.

Group	**Sample Size**	x_1	x_2	x_3
Anxiety	114	2.92	1.67	0.72
Psychopathy	32	3.81	1.84	0.81

(i) Obtain a linear discriminant function for classifying the data. **4**

(ii) Suppose a measurement on another subject yields $x_1 = 2.5$, $x_2 = 1.7$, and $x_3 = 0.88$. Would you classify it as Anxiety or Psychopathy? **3**

(iii) If among 114 subjects of Anxiety 78 are correctly classified and among all Psychopathy subjects, 20 subjects are misclassified. Then construct the confusion matrix and hence calculate APER. What conclusion will you make about this classification? **3**

First Tutorial **Time: 1:30**

Course Code: STAT-403 **Course Title: Design of Experiments**

1. Explain the term (i) Treatment (ii) Experimental unit (iii) Yield (iv) Block (v) Experimental error 5
2. Describe the concept of all basic principles of Design of experiment. Write down the purpose of them. 5
3. To compare the four mixtures, three different samples of propellant are prepared from each mixture and readied for testing. Each of three investigators is randomly assigned one sample of each of the four mixtures and asked to measure the propellant thrust. These data are summarized next. 10

Mixtures	Investigator		
	1	2	3
1	2,340	2,355	2,362
2	2,658	2,650	2,665
3	2,449	2,458	2560
4	2,403	2,410	2,418

i. Identify the blocks and treatments for this experimental design.

ii. Indicate the method of randomization.

iii. Write down the name of appropriate design for this data. Explain the reason for your choice.

iv. Conduct the ANOVA table for the data set when the observation for 3rd mixtures and 2nd investigator is missing.

2nd Tutorial

Course Title: Design and Analysis of Experiment II

Course code: STAT-404

Answer any one from 3 and 4

1. Define Concomitant variable with example. Write down the difference between analysis of variance and analysis of covariance. 5
2. A feeding trial experiment was conducted to compare the effects of two different feeds on the weight gain of goats in a firm. Feeds were given to selected goats for 3 months and gain weight are recorded. Tabulated result s of the experiment showing initial weight X (in kg) and gain in weight Y (in kg) 8

	F_1		F_2	
	x	y	x	y
	5.5	1	7.5	2.5
	6.5	2	5	1.5
	4.5	1	6.5	1.8
	7.5	1.6	6	2.0
Total	24	5.6	25	7.8

   i. Write down the appropriate ANCOVA model for this data. Justify the reason of your choice.
   ii. Complete the ANCOVA table.

3. Define variance component analysis. Write random effect model and its assumption for two-way Classification with single observation per cell. Conduct ANOVA table and estimate the variance components. Find the variance of the estimates. 7
4. The management of a poultry farm collected 10 varieties of dry concentrate to chicks of 7 different ages. Each concentrate was continued up to the age 45 days of chicks concentrate to the chicks and 5 varieties of dry concentrate and 4 different ages are randomly selected. The concentrates and different ages were given and then the weights of chicks were recorded. On the basis of this information complete the following table 7

SV	df	SS	E(MS)	F
Age	-	5.41	-	-
Concentrate	-	2.17	-	-
Error	-	0.55	-	
Total	-	-		

3rd Tutorial

Time: 50 Minutes **Total Marks:20**

1. Define factorial experiment. Distinguish between symmetrical and asymmetrical factorial experiments. Explain the algorithm for sign table to calculate different component sum squares in a factorial experiment in a RBD with 7 blocks. Also present the ANOVA table.

2. An experiment was conducted to see the effect of urea (N) and phosphate (P) fertilizers on the yield of a certain variety of rice: N and P were both at 2 levels (0 and 1).

Block							
I		II		III		IV	
I	n	n	I	I	n	np	n
22	30	28	24	25	30	32	31
P	np	P	np	p	np	p	I
40	30	36	36	35	30	39	30

   i. Write down the appropriate model for this dataset. Justify your answer.
   ii. Conduct the ANOVA table.

3. Define an incomplete block design. When incomplete block design becomes balanced. For a BIBD with usual parameters, show that

   i. $\lambda(v-1) = r(k-1)$

4. Describe the procedure of analysis of data obtained from a BIB design. Construct a layout plan for a BIB design having parameters b=v =7, r = k = 4, $\lambda = 2$.

CamScanner

4^{th} Tutorial

Time: 50 Minutes **Total Marks:20**

1. Discuss the block consists of 2^4-factorial experiment if ABCD and AC interactions are simultaneously confounded in the same replication. Discuss the procedure of analysis of data to test the hypothesis.

2. Define Split plot design. Discuss how does it differ from confounded design. Discuss the scopes and application of Split plot design. Explain why there are two types of error in SPD?

3. Imagine a manufacturing plant that produces a metal alloy used in various industrial applications. The goal is to optimize the alloy composition for strength, while also considering the impact of different heat treatment processes. Identify the main treatment and subplot treatments. Write down the appropriate model for this scenario. Describe different steps of analyzing the data of a split plot design.

CamScanner

Department of Statistics

Part IV B.Sc. (Honors) 1st Tutorial Examination, 2023

Course Title: Sampling Technique II

Course # 404

Time 45 Minutes Total marks 10

Q1.[1+2+4+3] a) What is an example of two-stage cluster sampling?

b) Describe a situation where two stage sampling is appropriate instead of one stage cluster sampling

c) Suggest an unbiased estimator of total of two-stage cluster sampling with unequal first-stage units. Verify your answer. How would you estimate the sampling variability of the estimator?

d) Find the best possible first stage units for two- stage sampling suggesting reasonable cost of the survey.

CS CamScanner

Jahangirnagar University

Department of Statistics

Part IV B.Sc. (Honors) (2nd) Tutorial Examination, 2022

Course Title: Sampling Technique II

Course # 404

Time 45 Minutes | Total marks 10

Answer the following questions.

Q1.a) In what survey situation would it be more appropriate to use probability proportion to size (PPS) sampling? Explain, in brief

b) Describe rejection method of selecting a sample by PPS with replacement sampling

c) Describe Sen-Midzuno sampling procedure. Show that, under this sampling $\hat{Y} = \sum_{i=1}^{n} \frac{y_i}{\pi_i}$

, π_i is the first order inclusion probability is an unbiased estimator of total

d) Reconsider Q1. c). Obtain an expression of sampling error

Q2. a) Find the gain in efficiency due to sampling with varying probability over simple random sampling with replacement (SRSWR)

CS CamScanner

Department of Statistics

Jahangirnagar University

Part IV B. Sc (Honors) First Tutorial Examination – 2022

Course No. : Stat- 405

Course Name: Data Mining

Answer the following questions. Each question carries equal marks.

Q1. Why naïve Bayesian classification is called "naïve"? Briefly outline the major ideas and steps of naïve Bayesian classification?

Q2. The following result (Table 1) was found for test data after applying the *k*-Nearest Neighbor (*k*NN) algorithm to the atmospheric data from a region of Bangladesh to classify the rainfall (RAN) [No Rain and Trace (NRT), Light Rain (LTR), Moderate and High Rain (MHR)] based on Temperature (TEM), Dew Point Temperature (DPT), Wind Speed (WIS), Humidity (HUM), and Sea Level Pressure (SLP) for the optimal value of $k=9$ and Seventy percent observations were used as training data and the rest of data as test data.

Table 1: Confusion matrix for the test data.

		Predicted		
	Category	LTR	MHR	NRT
Actual	LTR	65	6	11
	MHR	12	53	5
	NRT	8	0	73

(i) What is the actual number of observations?

(ii) Find the prediction accuracy rate, error rate for test data.

(iii) Obtain the value of Sensitivity, Specificity, and F_1-score for each category LTR, MHR, NRT.

Q3. Apply *k*-prototype clustering algorithm to find the cluster solution for $k = 2$ for the following data on $X_1, X_2, \ldots, X_5$. Use the ID-7 and ID-11 as initial cluster prototypes.

ID	X_1	X_2	X_3	X_4	X_5
1	13	14.5	5.7	F	P
2	12.2	10.3	5.8	M	C
3	13.3	13.9	5.7	F	P
4	13.3	15	6	F	P
5	12.7	12	5.8	M	C
6	13.4	15	5.7	F	P
7	12	10.7	5.7	M	C
8	12.3	10	5.8	M	C
9	12	9.5	5.8	M	C
10	12.7	11	5.8	M	C
11	13.7	15.2	5.8	F	P

Q4. Apply kNN algorithm to classify the item with information Sepal Length: 5.0, Sepal Width: 3.6, Petal Length: 1.4, and Petal Width: 0.3 based on the following training data for $k=5$. The

Department of Statistics
Jahangirnagar University
Part IV B. Sc (Honors) Second Tutorial Examination – 2022
Course No. : Stat- 405
Course Name: Data Mining

Answer the following questions. Each question carries equal marks.

Time: 1.5 Hours

Q1. What is meant by *k*NN? What are the different steps of *k*NN to identify the new objects? How can you find the optimal values of *k* for *k*NN?

Q2. What is meant by *k*-prototype clustering? What are the steps of *k*-prototype clustering algorithm? Also, describe the different steps to find the optimal value of *k* for *k*- prototype clustering method.

Q3. Write down the algorithm of Random Forest for Regression or Classification. Explain the following term Information Gain, Gain Ratio, Gini Index. Which measure is used as an attribute selection measure at ID3, C4.5, and Classification and Regression Tree model?

Q4. The following result found for training data after applying the CART algorithm to the atmospheric data from a region of Bangladesh to classify the rainfall (RAN) [No Rain and Trace (NRT), Light Rain (LTR), Moderate and High Rain (MHR)] based on Temperature (TEM), Dew Point Temperature (DPT), Wind Speed (WIS), Humidity (HUM), and Sea Level Pressure (SLP) and Seventy five percent observation used as training data and the rest of data as test data. *Explain the following results.* Also, find the F_1 score.

```
R-Output:
> confusionMatrix(as.factor(TR$RAN),PRR)
Confusion Matrix and Statistics
              Reference
Prediction LTR MHR NRT
       LTR 142  24  20
       MHR  35  86   1
       NRT  27   5 150
Overall Statistics
                   Accuracy : 0.7714
                     95% CI : (0.7316, 0.8079)
        No Information Rate : 0.4163
        P-Value [Acc > NIR] : <2e-16
                      Kappa : 0.6505
 Mcnemar's Test P-Value : 0.1239
Statistics by Class:
                     Class: LTR Class: MHR Class: NRT
Sensitivity              0.6961     0.7478     0.8772
Specificity              0.8462     0.9040     0.8997
Pos Pred Value           0.7634     0.7049     0.8242
Neg Pred Value           0.7961     0.9212     0.9318
Prevalence               0.4163     0.2347     0.3490
Detection Rate           0.2898     0.1755     0.3061
Detection Prevalence     0.3796     0.2490     0.3714
Balanced Accuracy        0.7711     0.8259     0.8884
```

Department of Statistics
Jahangirnagar University
Course name: Biostatistics and Epidemiology
Course Code: STAT-406 Quiz no: 01 Time: 01 hour

1. *Lifetimes (in hours) of particular insect larvae area modelled as observations of a random variable T with hazard function*

$$h_T(t) = \beta t^2$$

where $\beta > 0$ is an unknown parameter. A sample of n larvae are observed and survival times $t_i; :::; t_n$ recorded along with an indicator variable d_i, $i = 1; :::; n$ taking the value 1 if t_i is an observed death and 0 if observation i is censored at t_i.

(a) Find the density and survivor functions for T and hence write down an expression for the likelihood $L(\beta)$ for these data. *[8 MARKS]*

(b) Find an expression for the maximum likelihood estimator $\hat{\beta}$ of β, and calculate it for the data in the table below. *[8 MARKS]*

Observation (i)	1	2	3	4	5	6	7
Recorded time (t_i)	1	1	1	2	2	2	3
Observed death (d_i)	1	1	1	1	1	1	0

(c) If $\beta = 0{:}3$, calculate:

(i) the probability that a larva survives longer than 3 hours, *[2 MARKS]*

(ii) the time by which 90% of larvae have died. *[2 MARKS]*

Department of Statistics
Jahangirnagar University
Course name: Biostatistics and Epidemiology
Course Code: STAT-406 **Quiz no:** 02 **Time:** 01 hour

1. A study was undertaken to assess the effect of a new prison education programme on the rate of reoffending amongst released prisoners. Eight prisoners were followed up, four of whom had undertaken the education programme and four who had not. A Cox proportional hazards model was proposed for the data, where the hazard functions, h_E in the education group and $h_{\bar{E}}$ in the non-education groups, are related by

$$h_E(t) = exp(\beta)\, h_{\bar{E}}(t)$$

The following data were collected on the time (in months) between release from prison and committing a further criminal offence, a + indicating a right-censored observation.

Education group:	15+	28+	35	36+
Non-education group:	8	27	36+	36+

(a) Show that the partial log-likelihood for these data is given by

$$\boldsymbol{l(\beta) = \beta - 3\, log\left(1 + e^{\beta}\right) - log(24)}$$

and hence show that the maximum (partial) likelihood estimate of $\boldsymbol{\beta}$ is $\widehat{\boldsymbol{\beta}} = -0.693$ and $s.e.\left(\widehat{\boldsymbol{\beta}}\right) = 1.225.$ **[5 MARKS]**

(b) Calculate a 95% confidence interval for $\boldsymbol{\beta}$. What conclusions do you draw about the effectiveness of the education programme? **[4 MARKS]**

(c) The prison governors ask for a recommendation between:

A: Discontinue the new education programme

B: Extend the new education programme to all prisoners

C: Extend the study, collecting further data in both groups

Which of these would you recommend, and why? **[1 MARK]**

CS CamScanner

Jahangirnagar University

Course name: Biostatistics and Epidemiology

Course Code: STAT-406 **Quiz no:** 03 **Time:** 01 hour

1. Define and distinguish between the measures of disease frequency: prevalence and incidence. Provide examples of situations where each measure is more appropriate, and discuss their strengths and limitations. **5**
2. Compare and contrast the strengths and weaknesses of cohort and case-control study designs. Provide examples of research questions that are best addressed by each design and discuss the concept of temporality in epidemiological studies. **5**

4. In a rural county with 2000 children within preschool age, there have been 15 new cases of leukemia within 10 years. Compute the Incidence with 95% CI. **5**

3. In the study, 100 men with high-fat diets are compared with 100 men who are on a low-fat diet. Both groups start at age 65 and are followed for 10 years. During the follow-up period, 10 men in the high-fat intake group are diagnosed with prostate cancer and 5 men in the low-fat intake group develop prostate cancer. Compute 95% CI for incidence densities **5**

CamScanner

Department of Statistics

Jahangirnagar University

Course name: Biostatistics and Epidemiology

Course Code: STAT-406 **Quiz no:** 04 **Time:** 40 min

1. Suppose that T is a random variable representing survival time and that T has p.d.f. $f_T(t)$ hazard function $h_T(t)$and survival function $S_T(t)$

a. Write down the expression which relates $h_T(t)$ to $f_T(t)$ and $S_T(t)$ and hence show that

$$S_T(t) = \exp\left(-\int_0^t h_T(t)dt\right).$$

The operating time until failure (in hours) of particular LEDs are modelled as observations of a random variable T with hazard function,

$$h_T(t) = \beta t^{-\frac{1}{2}}$$

where $\beta > 0$ is a parameter of the model.

b. Find the density and survival functions for T.

c. If $\beta =$ 0.02, calculate:

i. The probability that an LED operates longer than 400 hours before failure,

ii. The time by which 90% of LEDs have failed.

2. In a study investigating an aggressive form of cancer, times (in months) were recorded between initial diagnosis and death for each patient. The following data were the recorded times in the study for 12 patients, a + indicating a right-censored observation.

1+ 32+ 5 13+ 7+ 2 3+ 24+ 14+ 2 31 14

a. Calculate the Kaplan-Meier estimate for the survival function of the random variable representing time to relapse. Sketch the estimate on a suitable set of axes (a very accurate sketch is not required, but you should label axes and show any points of discontinuity clearly).

b. Write down the estimate of the probability of survival for 2 years, and use Greenwood's formula to calculate its standard error. Hence calculate a 95% confidence interval for this probability.

CamScanner

Jahangirnagar University
Department of Statistics
Part IV B.Sc. (Honors) Examination - 2022
Course No.: Stat - 407
Course Name: Advanced Demography
1st Tutorial -25.07.2023

Time: 50 minutes | Marks: 10

N.B. Answer all the following questions.

1. Draw the demographic transition graph using five stages and explain the basic differences between 2nd stage and 3rd stage. **2.5**
2. Suppose a couple has 3 children (2 sons and 1 daughter) with wife aged is 48. What is the value of TFR, GRR, and NRR with explanations? **2.5**
3. Why growth rate varies among three projected methods such as linear, geometric, and exponential. Which method can be appropriate for projecting population in respect of Bangladesh and why? **2.5**
4. Why Bongaarts used the only four variables in his developed model, explain it. And write down the indicated model. Estimate the four indices with examples based on the given values for example, TFR=2.7, TMFR=4.2, u=0.581, e=0.88, i=5.8, and TA=0.18. **2.5**

Good Luck!

...ngh nagar University

Department of Statistics

Part IV B.Sc. (Honors) Examination – 2022

Course No.: Stat – 407

Course Name: Advanced Demography

2nd Tutorial - 09.11.2023

Time: 45 minutes **Marks: 10**

N.B. Answer all the following questions.

1. Define parity data and children ever born data. Why Gompertz model is applied to calculate age specific fertility rate instead of normal way? Explain Cumulative ASFR 2.8897 with age group 30-34. **5.0**
2. Based on the given table calculate the **female survivorship probability when age 40** including comments. **5.0**

Age group of respondents	Mother alive	Mother dead	Unknown maternal orphan-hood status	No. of children born in past years b(i)
15-19	5540	450	9	135
20-24	3990	540	5	410
25-29	2885	770	7	485
30-34	1905	850	6	320
35-39	1650	1235	9	260
40-44	1030	1270	4	95
45-49	850	1550	1	50
50-54	370	1240	4	

Good Luck!

Jahangirnagar University

Department of Statistics

1st Tutorial, Course Title: Stochastic Process, Course Code: STAT-408

Marks: 10 **Time: 50 Minutes**

---

1. Define Stochastics Process, Recurrent and Transient State of a Markov Chain, and Periodicity. Why Stationarity is important in Stochastic Process? **2**

2. State and Prove the First Entrance Decomposition Formula. **3**

3. On any given day Gary is either cheerful (C), so-so (S), or glum (G). If he is cheerful today, then he will be C, or S tomorrow with respective probabilities 0.5, 0.4. If he is feeling so-so today, then he will be C or G tomorrow with probabilities 0.3, 0.3. If he is glum today, then he will be S or G tomorrow with probabilities 0.3, 0.5.
   (a) Construct the Transition Probability Matrix. **1**
   (b) Draw the Transition Probability Graph. **1**
   (c) What is the probability of being cheerful on Tuesday given that Gary was glum on Sunday? **1**
   (d) Determine the ergodic states of the Markov Chain. **2**

CS CamScanner

Jahangirnagar University

Department of Statistics and Data Science

2nd Tutorial, Course Title: Stochastic Process, Course Code: STAT-408

Marks: 10 **Time: 45 Minutes**

1. Define Counting Process, Poisson Process, Compound Poisson Process, Renewal Process, Renewal Reward Process, and Stopping Time. **3**

2. Prove that with probability 1, $\frac{N(t)}{t} \to \frac{1}{\mu}$ as $t \to \infty$. **3**

3. There are two types of claims that are made to an insurance company. Let $N_i(t)$ denote the number of types i claims made by time t, and suppose that $\{N_1(t), t \geq 0\}$ and $\{N_2(t), t \geq 0\}$ are independent Poisson Processes with rates $\lambda_1 = 10$ and $\lambda_2 = 1$. The amount of successive type 1 claims are independent exponential random variables with mean \$1000 whereas the amount from type 2 claims are independent exponential random variables with mean \$5000. A claim for \$4000 has just been received; What is the probability it is a type 1 claim? **4**

CamScanner

Jahangirnagar University

Department of Statistics and Data Science

3rd Tutorial, Course Title: Stochastic Process, Course Code: STAT-408

Marks: 10 **Time: 45 Minutes**

---

1. Define *Queuing Process*, Mention assumptions of a queuing process. Derive the distribution of a Single Server Exponential Queuing System $M/M/1$ Having Finite Capacity. Also write down the Erlang loss formula. **4**

2. Define $M/G/1$ and $M/G/K$ Queuing model. Write down the relationship between $L, L_q, W,$ and W_q. **3**

3. Suppose that customers arrive at a Poisson rate of one per every 12 minutes, and that the service time is exponential at a rate of one service per 8 minutes. Find **3**
   i) Average number of customers in the super market
   ii) Average amount of time a customer spent in that queue to get into the super market.
   iii) If the arrival rate increases 20%, then what is the corresponding change in (i) and (ii)?

CamScanner

Tutorial 1	Jahangirnagar University Department of Statistics Savar, Dhaka-1342	Date: 01.06.2023

Course code: STAT-409 | Course Title: Bioinformatics | Full marks: 10

Answer the following questions.

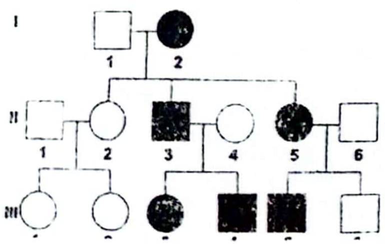
Q1. Define the following terms:
i) DNA, ii) Gene, ii) Chromosome, iv) Replication of DNA, v) Trait

Q2. Suppose, a SNP has two alleles. What are the possible genotypes you would have for this case? How does mutation related with replication of DNA? Explain.

Q3. Distinguish between a dominant allele from its recessive counterpart. Hence, relate these concepts with the genotypes obtained in Q2.

Q4. How would you define the penetrance probabilities in terms of the genotypes obatined in Q2.?

Q5.



Write the important characteristics of this pedigree chart.

CamScanner

Jahangirnagar University
Department of Statistics
Savar, Dhaka-1342

Date: 13.09.2023

Course code: STAT-409 | Course Title: Bioinformatics | Full marks: 10

Answer the following questions.

Q1. Define linage disequilibrium (LD) in the case of a pair of SNPs. Your answer should include the two SNPs with the allele frequencies and haplotype frequencies.

Suppose, there are three SNPs in a gene, and 0.25, 0.55, 0.88 are the disequilibrium coefficients for its pairwise level. What conclusions would you prefer for the gene?

Q2. How can you apply the penetrance function for defining the general genetic models? Explain for a SNP that follows HWE.

Q3. Express the genotype probabilities of the offspring when their parental genotypes are given. Consider the two SNP case with HWE.

CamScanner

Jahangirnagar University, Department of Statistics
Course No: 410 Course Name: Categorical Data Analysis,
Marks: 20, Time: 45 minutes, Date: 10-07-2023

1. Identify each variable as nominal, ordinal, or interval. **05**
a. UK political party preference (Labour, Conservative, Social Democrat).
b. Anxiety rating (none, mild, moderate, severe, very severe).
c. Patient survival (in number of months).
d. Clinic location (London, Boston, Madison, Rochester, Montreal)
e. Appraisal of company's inventory level (too low, about right, too high).

2. Consider the statement, "Please tell me whether or not you think it should be possible for a pregnant woman to obtain a legal abortion if she is married and does not want any more children." For the 1996 General Social Survey, conducted by the National Opinion Research Center, 842 replied "yes" and 982 replied "no." Let π denote the population proportion who would reply "yes." Test the hypothesis $\boldsymbol{H_0}$: $\boldsymbol{\pi = 0.50}$ using the score test, and Wald test. Interpret the results. **05**

3. Each of 100 multiple-choice questions on an exam has four possible answers, one of which is correct. For each question, a student guesses by selecting an answer randomly. **10**
a. Specify the distribution of $(\boldsymbol{n_1, n_2, n_3, n_4})$ where $\boldsymbol{n_j}$ is the number of times the student picked choice j. **b.** Find $\boldsymbol{E(n_j)}$, $\boldsymbol{and\ var(n_j)}$.

CS CamScanner

Department of Statistics
Tutorial Examination 2022
Course Title: *Categorical Data Analysis*
Full Marks: 40Course Code: STAT-410
Time: 1.30 Hours

All question carries equal marks.

Q1. Suppose you have a dataset following a Binomial distribution. Two models are considered: Model A with one parameter $_1$ and Model B with two parameters $_1$ *and* $_2$ .Perform a Likelihood Ratio Test to determine if adding the second parameter in Model B significantly improves the fit compared to Model A.

Q2. Consider a study investigating the relationship between the number of hours student study per day (X), and the probability of students G.P.A (Y). The data is collected from 100 students in a year under varying different conditions, and the response variable is binary (3.5 and above, below 3.5). State the appropriate probability distribution and link function for modeling the relationship between X and the probability of G.P.A (Y). Justify your choice.

Q3. Define and explain the concept of a saturated and unsaturated model. Provide an example to illustrate its application in the context of log linear modeling.

Q4. Formulate the logistic regression model for binary matched pair data. Define the key components of the model, including the dependent variable, independent variables, and parameters. Explain how this model can capture the relationship between the treatment and the binary outcome.

Q5. Explain different types of response variables used in modeling regression type model in the context of real-life scenario.

Q6. A sample of school girls are investigated and by whether using menstrual instruments are safe enough. Construct and interpret a 95% confidence interval for the population odds ratio. The data are as follows:

Instruments	Safe	Not Safe
Reuseable pad	10	18
Tissues	14	08
Cloth	13	20

Jahangirnagar University
Department of Statistics
Savar, Dhaka-1342

Date: 10.12.2023

Course code: STAT-409 **Course Title: Bioinformatics** Full marks: 10

Answer the following questions.

Q1. Mention at least three names for each type of the following databases.
(i) Nucleic Acid Sequence Database, (ii) Protein sequence Database, (iii) Genome Database.

Q2. State three important features for one database from each types that you have mentioned in (i), (ii) and (iii) of Q1.

Q3. Distinguish between, (i) Orthologous and Paralogous genes, (ii) Similarity and Homology of sequences, (iii) Global and Local alignment

Q4. What is GWAS? What are the impotance of it in bioinformatical studies?

engine

DEPARTMENT OF STATISTICS
JAHANGIRNAGAR UNIVERSITY
Savar, Dhaka-1342

Tutorial Examination, 2023

Student's Name: Tanima Jannat Oishi
Class Roll: 85 Session: 2018-2019
Course No.: STAT-401 Course Title: Statistical inference - II
Tutorial Exam. No.: 01 Date: 17-08-23

Signature of Examiner

Answer from here

Question

1. Consider following data: 170, 175, 180, 185. Obtain jack-knife estimate of mean. Test weather the estimator is unbiased or not. Find 95% CI for the jack knife estimator.

2. Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\theta, 1)$. Find pitman location estimator of θ and also show than, $\bar{X}_n$ is an unbiased estimator of θ.

3. Write down the pdf of truncated normal distribution truncated at range (a, b) and estimate the parameters of the distribution using MLE method.