Jahangirnagar University
Department of Statistics and Data Science
Class Test on Statistical Inference II

Course Code: STAT 401   Time: 60 min   Total Mark: 20

1.	Explain the concept of Fisher Information with usual notation. Let $X_1, X_2, \ldots, X_n$ be n independent random observations from the exponential distribution with mean $\frac{1}{\theta}$. Calculate the fisher's information under usual notation.	5
2.	Explain the concept of M-estimator with its properties. Show that the maximum likelihood estimator is a special case of an M-estimator.	5
3.	Explain elaborately the role of the Sensitivity Curve and Influence Function on Robust Statistics. For finite samples estimate the breakdown point of mean and median using the usual notation.	5
4.	Explain the concept of Jackknife and Bootstrap sampling with the limitations of using these resampling techniques.	5

Determine the MLE of the unknown parameter λ.

5) In a random sample of 85 automobile crankshaft bearings, 10 have a surface finish that is rougher than the specific allow. Construct a 95% two-sided confidence interval for p.   4

Jahangirnagar University
Department of Statistics and Data Science

Course Code: STAT 401   **Course Title: Statistical Inference-II**   **Time: 50 min**

1) Distinguish between probability distribution and Sampling distribution with examples.   3

2) What is statistical inference? What are the different types of statistical inference? What are the different methods of finding point estimators, explain one of them.   4

3) What is the necessity of hypothesis testing? Elaborate the procedure of Hypothesis testing and explain the types of errors can produce a test.   4

4) Let $X_1, X_2, \ldots, X_n$ be a discrete random sample from the Poisson (λ), $\lambda > 0$ distribution, with parametric space $\Theta = \lambda$.   5

$$f(X_j; \lambda = \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}, \lambda > 0$$

Determine the MLE of the unknown parameter λ.

5) In a random sample of 85 automobile crankshaft bearings, 10 have a surface finish that is rougher than the specific allow. Construct a 95% two-sided confidence interval for p.   4

Course Name: Bayesian Inference   Marks: 10
Course Code: STAT- 402   Time: 45 minutes
First Tutorial   Date: 03/10/2024

1. Suppose you have a Beta $(5, 5)$ prior distribution on the probability θ that a coin will yield a 'head' when spun in a specified manner. The coin is independently spun 20 times, and 'heads' appear fewer than 4 times. Calculate the posterior density along with its posterior mean.   [5]

2. A random sample of n students is drawn from a large population, and their weights are measured. The average weight of the n sampled students is 100 pounds. Assume the weights in the population are normally distributed with unknown mean and known standard deviation 30 pounds. Suppose the prior distribution is normal with mean 150 and standard deviation 20. Derive the posterior distribution and calculate the 95% posterior interval when $n = 20$.   [5]

Course Name: Bayesian Inference
Course Code: STAT- 402
Second Tutorial

Marks: 10
Time: 40 minutes
Date: 15/12/2024

1. Assume a conjugate prior for the two parameter univariate normal sampling model where,

$$\sigma^2 \sim INV - \chi^2(v_0, \sigma_0^2)$$

Derive the joint posterior distribution $p(\mu, \sigma^2 | y)$. [5]

2. Comparison of two multinomial observations: on September 25, 1988, the evening of a presidential campaign debate, ABC News conducted a survey of registered voters in the United States; 639 persons were polled before the debate, and 639 different persons were polled after. The results are displayed in Table:

Survey	Bush	Dukakis	No opinion/other	Total
pre-debate	294	307	38	639
post-debate	288	332	19	639

Assume the surveys are independent simple random samples from the population of registered voters. Model the data with two different multinomial distributions. For $j = 1, 2$, let α_j be the proportion of voters who preferred Bush, out of those who had a preference for either Bush or Dukakis at the time of survey j. What is the posterior probability that there was a shift toward Bush? Interpret your results. [5]

Time: 60 minutes   **Full marks: 20**

1. What is a multivariate normal distribution? How can you assess the assumption of multivariate normality? Let $X_1, X_2, \ldots, X_n$ are iid $N_p(\mu, \Sigma)$. Find the maximum likelihood estimate (MLE) of μ and Σ.

2. If X_1, X_2 and X_3 are jointly normal with quadratic form

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

   (i) Find the mean vector μ and covariance matrix Σ

   (ii) Find the conditional distribution of X_1 given $2X_2 + 3X_3$.

   (iii) Are X_1 and $2X_2 + 3X_3$ independently distributed? Explain.

Time: 60 minutes   Full marks: 20

The Madison, Wisconsin, police department regularly monitors many of its activities as part of an ongoing quality improvement program. Table 5.8 gives the data on five different kinds of overtime hours. Each observation represents a total of 16 pay periods or about half a year. This data is available in the file **T5-8.dat**.

Table 1: Five types of overtime hours for the Madison, Wisconsin, Police department

Legal Appearances Hours x_1	Extraordinary Event Hours x_2	Holdover Hours x_3	Compensatory overtime allowed x_4	Meeting Hours x_5
3387	2200	1181	14861	236
3109	875	3532	11367	310
⋮	⋮	⋮	⋮	⋮
3516+**X**	1223+**X**	1175+**X**	15078+**X**	161+**X**

X is the last two digits of your class roll number

1. Evaluate T^2 of the five variables $(x_1, x_2, \ldots, x_5)$ for testing
$H_0 : \mu' = \begin{bmatrix} 3500 & 1400 & 2600 & 13500 & 800 \end{bmatrix}$.

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

3. Construct a T^2-chart for data in Table 1 (**T5-8.dat**). Use $\alpha = 0.01$. Hence, comment.

4. At 1% level of significance, examine the equality of mean vectors of the observations on five types of overtime hours for the Madison, Wisconsin, police department and the Kansas City, Kansas Police Department (data on **T5_8_2.txt**). Assuming variances are equal.

5. Three psychological measurements x_1, x_2, and x_3 were taken on subjects in two neurotic groups ($\pi_1 =$ Anxiety, and $\pi_2 =$ Psychopathy). The sample means are given in the following table and the pooled covariance matrix is
$$S_p = \begin{bmatrix} 2.3 & 0.2 & 0.4 \\ 0.2 & 0.6 & 0.1 \\ 0.4 & 0.1 & 0.5 \end{bmatrix}.$$

Group	Sample Size	x_1	x_2	x_3
Anxiety	110	2.92	1.67	0.72
Psychopathy	35	3.81	1.84	0.81

   a) Test whether the profiles are parallel.
   b) Are the profiles coincident?
   c) Are the profiles level?

-Good Luck-

1. The following R output is obtained for conducting the factor analysis with 5 variables for $m = 1$ and $m = 2$ common factors, respectively.

```
Call:
factanal(x = x, factors = 1, method = "PCA", scale = T, c

Uniquenesses:
 Allied Dupont Carbide  Exxon  Texaco
 0.474  0.458  0.439   0.648  0.651

Loadings:
         Factor1
Allied   0.726
Dupont   0.736
Carbide  0.749
Exxon    0.593
Texaco   0.591

               Factor1
SS loadings     2.330
Proportion Var  0.466

Test of the hypothesis that 1 factor is sufficient.
The chi square statistic is 15.49 on 5 degrees of freedom
The p-value is 0.00847
```

```
Call:
factanal(x = x, factors = 2, method = "PCA", scale = T

Uniquenesses:
 Allied Dupont Carbide  Exxon  Texaco
 0.497  0.252  0.474   0.610  0.176

Loadings:
         Factor1 Factor2
Allied   0.601   0.378
Dupont   0.849   0.165
Carbide  0.643   0.336
Exxon    0.365   0.507
Texaco   0.207   0.884

               Factor1 Factor2
SS loadings     1.671   1.321
Proportion Var  0.334   0.264
Cumulative Var  0.334   0.598

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 followings:

(a) Check whether the 1 factor or the 2 factors are adequate for this model. Why?

For 1 factor; chi-square statistics with 5 degree of freedom and p value is 0.00847. A p value less than 0.05 indicates that the null hypothesis is rejected. Hence 1 factor is not sufficient for this model.

In factor-2; chi-square statistics is 0.58 with 1 degree of freedom and p value is 0.448. A p value greater than 0.05 indicates that the null hypothesis cannot be rejected. Hence factor 2 are sufficient for this model.

(b) Find the matrix of specific variances from the appropriate factor model. Therefore, identify the most important variable that perfectly fits into this factors model. Name the variable that doesn't match well in this factors model as well. Explain your findings.

(c) Find the estimated communalities. What proportion of the total population variance is explained by the first common factors from the adequate factor model?

(d) How can you estimate the factor scores?

2. 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
```

(a) List all the principal components with their variances.

(b) If the variances of length, width, and height are 0.0111, 0.0064, and 0.0060, respectively, then find and interpret the maximum correlation between the 1st principal components and the original variable.

(c) Draw the scree plot and decide how many principal components to retain in this study?

1. Set up a mathematical model for ANCOVA in RBD with one concomitants variable and   9
   discuss the analysis procedure of such data.
   i. Justify the impact of concomitant variable.
   ii. Check whether there are any differences in the effects of different levels of
   treatment.

2. In agricultural research station an experiment is conducted to study the productivity   11
   of 2 varieties of potato using nitrogen fertilizer. The agricultural plots for cultivation
   are found homogeneous in respect of fertility. The potato varieties are randomly
   allocated to different plots. But the amount of fertilizer used (x kg/plot) in different
   plots are not same. The production of potato (y kg) in different plots along with
   amount of fertilizer used are given below:

Plot	Potato 1		Potato 2	
	y	x	y	X
1	45	2	55	5
2	46	4	54	4
3	44	3	50	6

i. Write down the appropriate model for this data. Justify the reason of your choice.
ii. Complete the ANCOVA table.
iii. D Test whether the impact of concomitant variable is homogeneous or not for all varieties
   of potato.

1. Define factorial experiment. Write down the advantages of factorial experiment compared to single factor experiment. **4**

2. Construct the yates table to calculate different component sum squares in a 3^2 factorial experiment in a RBD with 7 blocks. Also present the ANOVA table. **6**

3. An experiment was conducted using three heterogeneous plots to see the effect of nitrogen N and irrigation I on the yield of a certain variety of rice. **10**

$$N = \begin{cases} 0 & \text{for 30 kg/ha} \\ 1 & \text{for 60 kg/ha} \\ 2 & \text{for 90 kg/ha} \end{cases} \qquad I = \begin{cases} 0 & \text{for low level} \\ 1 & \text{for moderate level} \\ 2 & \text{for high level} \end{cases}$$

Possible treatment combination and respective yield are given in the next table:

Plots					
I			II		
00	10	20	10	00	20
24	32	30	28	24	34
01	11	21	01	11	21
46	30	44	36	36	44
02	12	22	02	12	22
23	24	21	24	22	21

i. Write down the name of the design for this situation. Justify your answer.

ii. Write the mathematical model for this data.

iii. Estimate the components of ANOVA table. Construct the ANOVA table.

iv. Test which the treatment combination are similar.

1. Describe the procedure of intra-block analysis of data obtained from a BIB design. 10
Construct a layout plan for a BIB design having parameters $b = v = 13, r = k = 4, \lambda = 1$.

2. Define an incomplete block design. For a symmetric BIBD, prove that $\lambda = r_{ii'}$ where $r_{ii'} = \sum_{i=1}^{b} n_{ij} n_{i'j}; i \neq i' = 1(1)b$ and $\lambda = \sum_{i=1}^{b} n_{ij} n_{ij'}; j \neq j' = 1(1)v$ and also prove that $r - \lambda$ is perfect square.

3. Discuss the block consists of 2^3-factorial experiment if ABC and BC interactions 10 are simultaneously confounded in the same replication. Discuss the procedure for analyzing data to test the hypothesis.

4. Write down the advantages and disadvantages of confounding. Write down the scopes of confounding. Distinguish balanced and unbalanced confounding with examples.

Department of Statistics and Data Science
Jahangirnagar University
Part IV B. Sc (Honors) Examination - 2023
Course Code: Stat-405
Course Name: Data Mining

Time: 1 hour
Full Marks: 20
Date: 18/09/2024

Answer the following questions.

1. The following result is found for the test data after applying the k-nearest neighbor (kNN) algorithm to the atmospheric data from the Rajshahi 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 and seventy-five percent observations as training data and the rest of data as test data.   [05]

Table 1: Confusion matrix for the test data obtained from the kNN.

		Predicted		
	Category	LTR	MHR	NRT
Actual	LTR	2600	59	40
	MHR	181	148	0
	NRT	72	0	236

Table 2: Classification Report for the test data.

Name	Precision	Recall	F_1-score	Support
LTR	?	?	?	2699
MHR	?	?	?	329
NRT	?	?	?	?
Macro Average	?	?	0.765	3336
Weighted Average	0.887	?	?	3336
Accuracy →			?	3336

Complete the classification report based on the confusion matrix of test data and comment on your result.

2. Discuss the Minimal Cost-Complexity Pruning Algorithm to prune a decision tree.   [05]

3. What is meant by k-modes clustering? Write down the different steps of k-modes clustering. How can you obtain the optimal number of clusters for k-modes clustering?   [05]

4. The following results are produced using R from Bangladesh Demographic Health Survey-2022 training data based on some important variables Age of Respondents (Age: A1, A2, A3), Highest educational level of respondent (Education: Hi, NE, Pr, Se), Age of respondent at first birth (First_birth: LT18, GE18), Husbands education level (Partner _Education: Hi, NE, Pr, Se), Access to Healthcare (Healthcare: NProb, Prob) and Nutritional status (Nutrition: Malnutrition, Normal).   [05]

Department of Statistics and Data Science

```
> summary(TR$Nutrition)
Malnutrition        Normal
      2915            3064

> table(TR$Nutrition, TR$Age)
                A1      A2      A3
Malnutrition   427    1056    1432
Normal         613    1136    1315

> table(TR$Nutrition, TR$Education)
                Hi      NE      Pr      Se
Malnutrition   430     363     793    1329
Normal         369     495     894    1306

> table(TR$Nutrition, TR$First_birth)
               GE18     LT18
Malnutrition   1501     1414
Normal         1648     1416

> table(TR$Nutrition, TR$Partner_Education)
                Hi      NE      Pr      Se
Malnutrition   508     634     806     967
Normal         425     776     962     901

> table(TR$Nutrition, TR$Healthcare)
               NProb    Prob
Malnutrition   1178     1737
Normal         1160     1904
```

Using Naive Bayes classification predict nutritional status of the following test data from BDHS-2022. Finally, comment on your results.

Table 3: Test Data from BDHS-2022.

ID	Age	Education	First_birth	Partner_Education	Healthcare
1	A3	Hi	LT18	Se	NProb
2	A1	Se	GE18	NE	Prob
3	A2	Pr	LT18	Hi	Prob

[04]

[04]

04]

[04

End of exam questions.

Department of Statistics and Data Science
Jahangirnagar University
Part IV B. Sc (Honors) Examination - 2023
Course Code: Stat-405
Course Name: Data Mining

Time: 75 minutes
Full Marks: 20
Date: 27/11/2024

Answer the following questions.

1. Given that, $A(1,2)$, $B(1,3)$, $C(4,3)$, $D(5,4)$, $E(8,6)$, and $F(8,7)$ be the six points and the initial cluster centers are $C_1 = (1.5, 2.5)$, $C_2 = (4.5, 4.5)$, and $C_3 = (8.5, 6.5)$. Apply the Fuzzy C-means clustering algorithm with the maximum number of iterations 4 to find the cluster solution and find the Fuzzy Partitioning Coefficient. . **[06]**

2. What is meant by k-medoids clustering? Write down the different steps of the k-medoids clustering algorithm. What are its different advantages from other clustering techniques? **[04]**

3. Apply Self Organizing Map (SOM) to cluster the A, B, C, and D data points for an iteration. Assume that the initial learning rate is 0.5 and the number of clusters to be **[05]**

i	A	B	C	D
1	1	0	1	1
2	1	1	0	1
3	1	1	0	0
4	0	0	0	1

formed is 2. Consider the initial weight matrix

$$W = \begin{bmatrix} 0.2 & 0.9 \\ 0.4 & 0.7 \\ 0.6 & 0.5 \\ 0.8 & 0.3 \end{bmatrix}$$

4. The following contingency table summarizes supermarket transaction data (in thousand), where B refers to the transactions containing hot dogs, $\overline{B}$ refers to the transactions that do not contain hot dogs, A refers to the transactions containing hamburgers, and $\overline{A}$ refers to the transactions that do not contain hamburgers. **[05]**

	B	$\overline{B}$	Total
A	20	5	25
$\overline{A}$	10	15	25
Total	30	20	50

Obtain the value of Kappa, Imbalance Ratio, Laplace, Certainty factor, and Collective strength.

End of exam questions.

Department of Statistics and Data Science
Jahangirnagar University
Part IV B. Sc (Honors) Examination - 2023
Course Code: Stat-405
Course Name: Data Mining

Time: 75 minutes
Full Marks: 20
Date: 09/01/2025

Answer the following questions.

1. What is meant by Support Vector Machine? What are the different types of kernels used in support vector machine (SVM)? **[04]**

2. Write down the different steps of support vector machine (SVM) for classification. **[04]**

3. Define the different types of activation functions. Describe the backpropagation method of training an artificial neural network. **[04]**

4. Consider a grayscale image and a convolutional filter represented as follows: **[04]**

$$\text{Image} = \begin{bmatrix} 10 & 11 & 9 & 2 & 23 & 5 \\ 5 & 1 & 8 & 3 & 17 & 8 \\ 4 & 6 & 5 & 9 & 8 & 9 \\ 7 & 2 & 7 & 3 & 7 & 6 \\ 2 & 3 & 8 & 5 & 2 & 5 \\ 11 & 4 & 9 & 4 & 5 & 5 \end{bmatrix}$$

$$\text{Filter} = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

Find Max Pooling and Average Pooling Feature Map using a 2x2 window with stride 2 after convolution with stride 1 and ReLU.

5. What is text mining? What are the different types of Text mining techniques? What is lexicon-based sentiment analysis? Explain the methods to Compute Sentiment Scores, Lemmatization, Tokenization, Sentiment score, and VADER. **[04]**

End of exam questions.

Total marks: 20   Total Time: 60 Minutes

Answer the following questions.

Q1. a) What is meant by nominal rate of interest and force of interest?

b) Under usual notations, prove that

i) $i = \left[1 + \frac{i^{(m)}}{m}\right]^m - 1$ 　　　 ii) $v = \frac{1}{1+i}$ 　　　 iii) $i^{(n)} = n\left\{\left(\left[1 + \frac{i^{(m)}}{m}\right]^m\right)^{\frac{1}{n}} - 1\right\}$

c) Mention some practical applications of annuities and perpetuities

d) Under usual notations, show algebraically that

i) $a_{\overline{\infty}|} = \frac{1}{i}$ 　　　　　 ii) $a_{\overline{n}|} = \frac{(1+i)^n - 1}{i}$

iii) $s_{\overline{m+n}|} = s_{\overline{m}|} + (1+i)^n s_{\overline{n}|}$ 　　　 iv) $Ia_{\overline{n}|} = \frac{1 - (\frac{1+g}{1+i})^n}{i - g}$

e) What is a loan amortization schedule? Consider a loan which is being repaid by equal annual payments of 1 for n years. Construct an amortization schedule

f) A loan of L is to be repaid by sinking fund method over n years. Find the (equal) periodic payment of the sinking fund. Also, construct sinking fund schedule

Q2. a) Explain mathematically and verbally why the following expressions are true?

　　i)) $_nq_x = 1 - \frac{l_{x+n}}{l_x}$ 　　 ii) $_{m+n}p_x = (_mp_x)(_np_{x+m})$

b) What is the difference between net premium and office premium?

c) Draw a cash flow diagram of an n-year term insurance of face value 1 on (x). Derive the expression to find net single premium

d) An annuity on (x) provides k annually beginning at age $(x + n)$. Nothing is paid before n years. Annual premiums are payable for n years beginning at x. Draw a cash flow diagram and obtain the expression for annual premium, assuming limiting age ω

Jahangirnagar University
Department of Statistics
Part IV B.Sc. (Honors) Examination – 2023
Course No.: Stat – 407
Course Name: Mathematical Demography
1ˢᵗ Tutorial -24.10.2024

Time: 50 minutes   Marks: 10

N.B. Answer all the following questions.

1. Define age and sex composition with example. Generate hypothetical data and draw graph and explain it.   2.5
2. In which situations, you need to apply Myers index instead of Whipplex Index?   2.5
3. In what conditions stable population differ from stationary population? Please explain the theoretical concepts of Lotka's stable population.   2.5
4. Why do we need to project future population? Are there any specific conditions to apply different mathematical methods of projecting populations? If exists then explain with graphically too.   2.5

Good Luck!

Jahangirnagar University
Department of Statistics and Data Science
Part IV B.Sc. (Honors) Examination – 2023
Course No.: Stat – 407
Course Name: Mathematical Demography
2nd Tutorial - 03.12.2024

Time: 45 minutes   Marks: 10

N.B. Answer all the following questions.

1. When do you need to apply Bongaart's and Gompertz's models, and why? The approximate estimated values of 2023 in Bangladesh are TFR = 2.17, TMFR=4.2, Proportion of contraceptive use is 0.59, contraceptive use effectiveness is 0.91, Average Duration of Postpartum Amenorrhoea is 6, and TA rate is 0.30. Estimate 4 indices of the Bongarts model and find out which one is the most responsible estimate to increase TFR according to the Bongaarts model. Also, interpret the results.   5.0

2. Define census coverage. Estimate Birth rate and mortality level using two-census age distribution with indicating assumptions and using Brass Logit Transformation system and Preston and Coalse stable population model.   5.0

Good Luck!

Jahangirnagar University
Department of Statistics and Data Science
1ˢᵗ Tutorial, Course Title: Stochastic Process, Course Code: STAT-408

Marks: 10   Time: 50 Minutes

1. Define Stochastics Process. What are the different types of Stochastic Processes?   **2**

2. Define Brownian motion with its properties. Hence write down some of its applications   **2**

3. Let $\{X_n \geq 0, n \geq 0\}$ is a stochastic process and the state space are 0: No correction required and 1: Correction required. The signal of the system has been recorded at 10 successive hours instants, and the recorded sequence is: 0 0 0 1 1 0 1 0 1 1

    i)   Estimate the TPM function of the signal of the system.   **1**

    ii)   Draw the transition probability graph.   **1**

    iii)   What is the probability that the process will signal correction required at the 11ᵗʰ hour given that it signaled correction required at the 9ᵗʰ hour?   **1**

    iv)   Specify the classes of the Markov chain and determine whether they are transient or recurrent. **1**

    v)   On average, after every how many hours does the process signal no correction required?   **2**

(c) Does {N(t)} have stationary increments? Why or why not?

(d) Does it have independent increments? Why or why not?

(e) If next year starts off with three storms by time $t=2$, what is the conditional probability it is a good year?

Marks: 10   Time: 50 Minutes

1. When is a Counting Process said to be a Poisson Process? Derive the distribution of Poisson Process.   4

2. Define the Renewal Reward Process. Find the Second Moment and Variance of the Renewal Process.   3

3. In good years, storms occur according to a Poisson process with rate 2 per unit time, while in other years they occur according to a Poisson process with rate 4 per unit time. Suppose next year will be a good year with probability 0.4. Let $N(t)$ denote the number of storms during the first t time units of next year.   3

   (a) Find $P\{N(t)=n\}$.
   (b) Is $\{N(t)\}$ a Poisson process?
   (c) Does $\{N(t)\}$ have stationary increments? Why or why not?
   (d) Does it have independent increments? Why or why not?
   (e) If next year starts off with three storms by time $t=2$, what is the conditional probability it is a good year?

1. Write down the relationship between Birth and Death process. State and prove Kolmogorov's
Backward Equations. **3**

2. Define the Queueing Process with its elements. Derive the distribution of Erlang loss system for
finite capacity. **4**

3. A mechanic looks after 8 automatic machines, a machine breaks down, independently of others, in
accordance with a Poisson process, the average length of time for which a machine remains in
working order being 12 hours. The duration of time required for repair of a machine has an
exponential distribution with mean 1 hour. Find: **3**

   i) The probability that 3 or more machines will remain out of order at the same time.

   ii) The average number of machines in working order.

   iii) For what fraction of time, on the average, the machine will be idle?

Course code: STAT-409   Course Title: Bioinformatics   Full marks: ~~10~~ 20

Time: 1 hour

Answer the following questions.

Q1. Define the following terms:
  i) Locus, ii) Genotype, ii) Allele, iv) Diallelic, v) Phenotype

Q2. Suppose, a SNP has two alleles. How would you define the dominant and recessive alleles for that case? Explain.

Q3. What is the penetrance function? How would you define the penetrance probabilities in terms of the genotypes for a gene with single SNP? Explain.

Q4.
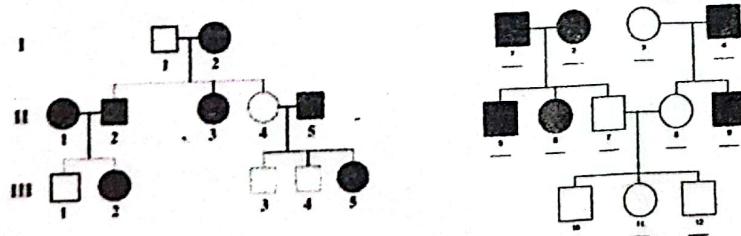
Compare the two pedigree charts.

Full Marks: 20

Time: 1 Hours

[Answer ALL the questions]

1. a) What is epistasis? What are the different methods for testing gene gene interaction? Describe one of them. **4**

   b) Define fixed and random effects models for meta-analyses. How fixed and random effects models for meta-analyses are applied in genetic association study? How can we evaluate the clinical validity of a test based on a single genetic marker? **5**

2. a) A study examines the association between *GeneY* and hypertension, with three genotypes: *bb, bB,* and *BB*. The data below shows the number of individuals with and without hypertension: **5**

Genotype	Hypertension (Affected)	No Hypertension (Not Affected)	Total
bb	x_i 30	x_o 70	100
bB	x_1 45	y_o 55	100
BB	70 x_1	30 x_o	100

   1. Calculate the odds of hypertension for each genotype and interpret the results.
   2. Find the odds ratios for *bB vs. bb* and *BB vs. bb*, then interpret these ratios in terms of hypertension risk.
   3. Compute the log odds ratios θ_1 and θ_2 for *bB vs. bb* and *BB vs. bb*. Interpret what these values imply about the relative risk of hypertension.
   4. Based on your findings, select the appropriate genetic model to describe the relationship between *GeneY* and hypertension.

   b) How does population stratification impact genetic association studies, and why is it crucial to detect and control for this form of confounding, especially in large-scale genome-wide association studies (GWAS)? What are the different methods to detect control for this form of confounding? Describe at least three methods. **6**

Best of Luck

Department of Statistics and Data Science
Jahangirnagar University
First Tutorial Examination
Course Title: Statistical Genetics and Bioinformatics
Course Code: STAT-409

Time: 1 Hours   Full Marks: 20

[Answer ALL the questions]

1. a) What are genetic markers, how do we locate them, and how their quality assessed? What are   4
   their real-world applications?

   b) Define SNP, and how does it differ from general polymorphism? Can you explain   4
   genotyping and what are the various genotyping methods, and how to select the appropriate
   genotyping method?

2. a) What is Hardy-Weinberg Equilibrium (HWE), and what factors can disrupt it? Under the   6
   general condition of HWE,

   $$P(A_1) = p_{11} + \frac{1}{2}p_{12} \text{ and } P(A_2) = p_{22} + \frac{1}{2}p_{12} \text{ where } p + q = 1$$

   find the relationship between genotype and allele frequency.

b) Linkage disequilibrium (LD) is an important concept in understanding the genetic 6
architecture of complex diseases like Type 2 diabetes (T2D). Given two loci, A and B, where
each has two alleles, A_1, A_2, B_1 and B_2 the following haplotype frequencies are observed in
a population: $P(A_1, B_1) = 0.35, P(A_1, B_2) = 0.15, P(A_2, A_1) = 0.15$ and $P(A_2 B_2) = 0.35$.
Using these values, answer the following:

1. Calculate the marginal allele frequencies of A_1, A_2, B_1 and B_2.
2. Compute the linkage disequilibrium coefficient D and explain its significance in
    terms of the association between loci A and B.
3. Determine the normalized LD measure D' and the squared correlation coefficient
    r^2, and interpret their biological implications in the context of genetic studies.

Best of Luck

Given Data : The haplotype frequencien arce

. $P(A_1 B_1) = 0.35$

$(A_1 B_2) = 0.15$

Jahangirnagar University, Department of Statistics
Course No: 411, Course Name: Financial Statistics,
Marks: 10, Time: 30 minutes, Date: 01-10-2024

1. What does the stock market mean when you say "bullish" or "bearish"? 05
   Differentiate between the primary and secondary markets. Briefly discuss
   the efficient-market hypothesis.

2. What is an ARCH model? How to model volatility with ARCH for financial 05
   time series forecasting

Jahangirnagar University, Department of Statistics
Course No: 411, Course Name: Financial Statistics,
Marks: 10, Time: 50 minutes, Date: 05-12-2024

1. Describe the Dow Theory. How Elliott Waves and Dow Theory Operate. Explain some of the common patterns of impulse and corrective waves. 05

2. Describe the bullish and bearish divergences in the MACD. Calculate and analyze the ATR for the given data: 05

Date	High	Low	Close	Close
Day-1	249.83	240.70	240.83	274.45
Day-2	246.21	240.56	244.50	279.7
Day-3	247.43	239.51	241.05	274.2
Day-4	242.79	232.34	238.77	272.1
Day-5	223.34	214.38	217.80	240.83
Day-6	183.60	180.25	180.80	244.5
Day-7	183.09	180.92	182.72	241.05
Day-8	185.85	182.05	185.17	238.77
Day-9	188.13	185.83	186.65	217.8
Day-10	249.83	240.70	240.83	180.8