UNIVERSITY OF THE PUNJAB

Part-II A/2016
Examination:- M.A./M.Sc.

Roll No. ....

Subject: Statistics **TIME ALLOWED: 3 hrs.**
PAPER: I (Statistical Inference) **MAX. MARKS: 100**

NOTE: Attempt any FOUR questions. All questions carry equal marks.

Q.1 a) Define consistency, write different types of consistency; write comparative note on these types. (10)

b) Let $X_1, X_2 \sim p(\lambda)$ i.e. distributed as Poisson with parameter λ; then prove that $T_1 = X_1 + X_2$ is sufficient statistic for λ and $T_2 = X_1 + 2X_2$ is not sufficient statistic for λ. (10)

c) Write and explain two methods for finding out the sufficient estimator. (05)

Q.2 a) Let $X_1, X_2, \ldots\ldots\ldots, X_n$ be a random sample from some population with finite mean and variance. Prove that $s^2 = \sum(X - \bar{X})^2/(n-1)$ is an unbiased estimator for population variance but $S = \sqrt{\sum(X - \bar{X})^2/(n-1)}$ is not, in general, unbiased for population standard deviation, hence prove that the bias, (12)

$$b(\sigma, s) = \sigma\left\{\sqrt{\frac{2}{n-1}}\Gamma(n/2)\left[\Gamma\left(\frac{n-1}{2}\right)\right]^{-1} - 1\right\}$$

b) Find the minimum variance unbiased estimators for the parameters of normal distribution, what are their variances? (08)

c) Define invariance. Mention one invariant estimator. Compare asymptotic unbiasedness with consistency. (05)

Q.3 a) A sample of size n_1 is to be drawn from $N(\mu_1, \sigma_1^2)$. A second sample of size n_2 is to be drown from $N(\mu_2, \sigma_2^2)$ what is MLE for $\alpha = \mu_1 - \mu_2$? Assuming that the total sample size n = $n_1 + n_2$ is fixed. How the n observations should be divided between the two populations in order to minimize the variance of $\hat{\alpha}$. (12)

b) A sample of size 'n' is drawn from Normal populations all of which have the same variance σ^2. The mean of the four populations are a+b+c, a+b-c, a-b+c and a-b-c, what are the MLEs for a, b, and c. (13)

Q.4 a) Define Moment estimator, write the rational of finding the moment estimators. (07)

PTO

b) Estimate the parameters of log normal distribution using moment method of estimation. (08)

c) Let $P(X=x)=\binom{x+\theta_2-1}{x}\theta_1^{\theta_2}(1-\theta_1)^x \qquad x=0,1,\ldots\ldots,\infty$ (10)

Find the moment estimators of θ_1 and θ_2

Q.5 a) Based on random sample of size n from Poisson distribution obtain minimum χ^2 estimator (MCSE) and MLE of θ and compare the two estimators. (06)

b) Compare the properties of Maximum likelihood estimator and OLS estimator, which is the better estimator? Give reasons. (09)

c) If X is a Poisson variate with parameter λ and if $g(\lambda)=\frac{1}{m!}\left(\frac{m+1}{\lambda_0}\right)^{m+1}\lambda^m e^{\frac{-(m+1)\lambda}{\lambda_0}}$ is prior density for unknown parameter λ, given a random sample of size n, obtain the Bayes' estimator for λ. (10)

Q.6 a) State and prove Neyman-Pearson's lemma. (12)

b) Explain the statistical method for the construction of confidence intervals. Also distinguish between confidence intervals and confidence region. (08)

c) Let x follow Bernoulli population, perform SPRT to test $H_0:\theta=1/3$ *against* $H_1:\theta=2/3$ (05)

Q.7 a) Define best critical region (BCR) and power of the test. What is the relationship between the two? (06)

b) Let $x_1, x_2, \ldots\ldots, x_n$ denote a random sample from a distribution that has p.d.f. $f(x_i)$ that is positive on only non-negative integers. It is desired to test the simple hypothesis $H_o:f(x)=e^{-1}/x!, \quad x=0,1,2,\ldots\ldots$ against alternative simple hypothesis $H_1:f(x)=(1/2)^{x+1} \quad x=0,1,2,\ldots\ldots$ Derive the expression for BCR (Best critical region). Consider the case of n=1 and k=1, k being any positive integer in the expression $\frac{L(\theta',x_1,x_2,\ldots\ldots,x_n)}{L(\theta'',x_1,x_2,\ldots\ldots,x_n)}\le k$ (12)

where $H_o:\theta=\theta', \quad H_1=\theta=\theta''$.

Find the power of the test for this combination of n and k when

(i) H_o is true (ii) H_1 is true

c) What do you mean by ASN (Average Sample Number) in which type of sampling it is used? Write the methods to find the ASN. (07)

UNIVERSITY OF THE PUNJAB

Part-II A/2016
Examination:- M.A./M.Sc.

Roll No. ....

Subject: Statistics **TIME ALLOWED: 3 hrs.**
PAPER: II (Regression Analysis and Econometrics) **MAX. MARKS: 100**

NOTE: Attempt any FOUR questions. All questions carry equal marks.

Q.1.a) Discuss the functions of Econometrics. Under what reasons an error term is introduced in Econometric models?

b) Show that estimated error variance for model

$Y_i = \alpha + \beta X_i + \epsilon_i$ *is* $\frac{1}{n-2}\left(\Sigma Y_i^2 - \hat{\alpha}\Sigma Y_i - \hat{\beta}\Sigma X_i Y_i\right)$ where $\hat{\alpha}$ and $\hat{\beta}$ are OLS estimator.

Also find its expected value.

Q.2.a) Consider the model $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots\ldots\ldots + \beta_k X_k + \epsilon$, Which fulfills the OLS assumptions. Develop the testing procedure for $H_0 : \beta_1 = \beta_{10}$, $\beta_2 = \beta_{20}, \ldots\ldots, \beta_k = \beta_{k0}$

b) Consider the model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$, with sample data, $n = 100$, $\Sigma x_1^2 = 30$, $\Sigma x_2^2 = 3$, $\Sigma x_1 x_2 = 0 = \Sigma x_1 y = 30$, $\Sigma x^2 y = 20$, $\Sigma y^2 = \frac{493}{3}$, Test the hypotheses $H_0 : 7\beta_1 - \beta_2 = 0$ *against* $H_1 : 7\beta_1 - \beta_2 \neq 0$

Q.3.a) What are the orthogonal polynomials? Discuss there uses in regression analysis.

b) Consider the models $\underline{Y} = X_1 \underline{\beta}_1 + X_2 \underline{\beta}_2 + \underline{\epsilon}$. suppose we have unbiased estimate $\underline{\beta}_1{}^*$ of $\underline{\beta}_1$. Obtain the least squares estimates of $\underline{\beta}_2$. Compare its sampling variance with that of unrestricted estimator of $\underline{\beta}_2$.

Q.4.a) State and prove Aitken Theorem.

b) Let $\underline{Y} = X \underline{\beta} + \underline{\epsilon}$ *Such that* $X's$ are stochastic regressors. State the necessary assumptions and find asymptotic variance of OLS estimator of β.

Q.5.a) Define Heteroskedasticity and discuss some measures to remove it.

b) Differentiate between collinearity and Multicollinearity. Discuss different indicators used for detection of multicollinearity.

Q.6.a) What is autocorrelation? How the parameters of a regression model are estimated in the presence of autocorrelation? Discuss

b) When and how instrumental variables are selected. Show that instrumental variables estimators are consistent.

Q.7. Consider the following model $y_{1t} = \beta y_{2t} + u_{1t}$ $\quad y_{2t} = \alpha_1 y_{1t} + \alpha_2 x_{1t} + \alpha_3 x_{2t} + u_{2t}$

(i) Show that OLS estimate of β is inconsistent estimate

(ii) Obtain consistent estimates of the structural parameters β and $\alpha's$, where possible, by appropriate method using the following calculations

$\Sigma x_1^2 = 1$, $\Sigma x_2^2 = 20$, $\Sigma x_1 x_2 = 0$, $\Sigma x_1 y_1 = 5$, $\Sigma x_2 y_1 = 40$, $\Sigma x_1 y_2 = 10$, $\Sigma x_2 y_2 = 20$

UNIVERSITY OF THE PUNJAB

Part-II A/2016
Examination:- M.A./M.Sc.

Roll No. ....

Subject: Statistics **TIME ALLOWED: 3 hrs.**
PAPER: III (Part-A) [Data Processing and Computer Programming] **MAX. MARKS: 75**

NOTE: Attempt any FOUR questions.

Q.1(a) What are the uses of following DOS commands, explain with example
FORMAT, MD, CHKDSK, COMMAND

(b) Differentiate between high level and low level languages; how they are different from middle level language.

(c) Differentiate between source program and object program. (10+5+4)

Q.2(a) How are the FORTRAN 77 statements coded using FORTRAN coding sheet? Also what is the purpose of putting a ' C' in column 1 of the Ftn coding sheet.

(b) Write the following expression in Fortran notation.

(i) $(4x^2+2y+xy+5)^3\sqrt{2(xy-6)^4}$ (ii) $e^x+\sin x+3.0\sqrt{yx}$

(iii) $\dfrac{zm(1+x)}{[1-(\frac{v}{c})^2]^{\frac{1}{2}}}$ (iv) $\dfrac{m}{[1-(\frac{v}{c})^2]^{\frac{1}{2}}}$

(c) Rewrite the following statements after correcting errors, if any:
(i) Read , (card(I , J) , I = 1 , 15 , J=1 , 10)
(ii) Write , ((J, I =1 , 10) , I , J =1 , 15)
(iii) Read Doly (I), I= 1 ,10
(iv) Write (* , *) X , Z

(d) Determine the output of the following programs:

```
C  First program
   J=5
   K=4
   L=K+2*J
   J=2*L+J/5
   K=K/5
   L=J+K+L
   WRITE(*,*) J, K,L
   STOP
   END
```

```
C  Second program
   A = 5.2
   B = 512.3
   W = 17.4
   WRITE(*,10) A, B, W
10 FORMAT(T10,F6.2,//,T10,F6.2,//,T10,F6.2)
C
   STOP
   END
```

(4+5+4+6)

Q.3(a) Write down the general form of the arithmetic IF and computed GOTO statements.

(b) Write a program in FORTRAN to Compute the total possible samples with and without replacement for n=5 and r = 3.

(c) Income Tax for individuals is computed in slab rates as follows:

Upto Rs. 100,000	Nil
Between Rs 100,000 and Rs. 150,000	10% of excess over Rs100,000
Between Rs 150,000 and Rs. 250,000	Rs 5000 + 25% of excess over Rs150,000
Above Rs. 250,000	Rs. 30,000 + 40% of excess over Rs250,000

Wriite a program in FTN 77 that reads amount and print the Income Tax due.

(d) Print the numbers I, A, J, K, B, C, L, M, N, D, E, and F on three lines so that the first line should contain one integer and one real number; the second line should contain two integers and two real numbers; and the third line should contain three integers and three real numbers. For each integer the field should be six; each real number should be printed in a 10 - column field with three decimal places. (4+4+8+3)

P.T.O.

Q.4(a) Define the functions of the following FORTRAN statements. Give examples at least two in each case;

(i) TYPE STATEMENT
(ii) DO STATEMENT
(iii) LOGICAL IF STATEMENT.

(b) Write a complete Fortran program to find the Standard deviation, coefficient of variation and coefficient of skewness for N real numbers.

(c) Suppose an examination is given to a class of exactly 41 students and their scores are read one per line. Write a complete Fortran program to find the class average, class median and count the number of students who scored above the average.

(6+6+7)

Q.5(a) What is an Array? What are the advantages of using Arrays?

(b) Describe the difference between a FUNCTION subprogram and a SUBROUTINE subprogram.

(c) Given a matrix A of size MXN. Write a program to find the sum of squares of the reciprocal of each row and store it in a one-dimensional array and print the largest sum.

(6+6+7)

Q.6(a) What is the use of following:

%f	%s
%c	\n
\t	\r
\b	*=
\?	\\

(b) Distinguish between a **switch** statement and an **else-if** statement.

(c) Write programs using switch statement and else-if statement to make a four function calculator.

(d) Define the fowolling statements with examples.

(i) For Loop,
(ii) Do - While Loop.

(4+4+8+3)

Q.7(a) What is the differnece between **#define Directive** and **#include Directive**?

(b) Write a recursive function to find binary equivalent of a number.

(c) Define structure; write a c-program which can keep records of a class of students holding the fields of roll numbers, names, class in which they study, blood group and year of enrolment. The program should also print the records in the order of their blood group.

d) Write a C-program that displays the sum of following series

1/2, 3/4, 7/8, 15/9 ....up to 50 values

by writing a program using 'functions'

(4+7+8)

UNIVERSITY OF THE PUNJAB

Part-II A/2016
Examination:- M.A./M.Sc.

Roll No. ....

Subject: Statistics **TIME ALLOWED: 3 hrs.**
PAPER: VI (i) [Statistical Quality Control] **MAX. MARKS: 100**

NOTE: Attempt any FOUR questions. All questions carry equal marks.

Q#1 (a)	Explain in brief three important postulates concerning the laws basic to control.	10
(b)	What are the advantages of statistical quality control?	15
Q#2 (a)	Samples of size $n = 5$ are taken from a manufacturing process every hour. A quality characteristic is measured, and $\bar{x}$ and R are computed for each sample. After 25 samples have been analyzed, we have $\sum_{i=1}^{25} \bar{x}_i = 662.50$ and $\sum_{i=1}^{25} R_i = 9.00$. The quality characteristic is normally distributed. 1. Find the control limits for the $\bar{x}$ and R charts. 2. Assume that both charts exhibit control. If the specifications are 26.40 ± 0.50, estimate the fraction non-conforming. 3. If the mean of the process were 26.40, what fraction non-conforming would result?	15
(b)	In order to meet Government regulations, the contained weight of a product must at least be equal to the labeled weight 98% of the time. Control charts for $\bar{x}$ *and* σ are maintained on the weight in ounces of the contents, using a subgroup size of 10, after 20 subgroups, $\sum \bar{x} = 731.4$ and $\sum \sigma = 9.16$ Compute 3-sigma control limits for $\bar{x}$ *and* σ and estimate the value of σ' assuming that the process is in statistical control. If the labeled weight is 36 oz, and assuming the process generator a normal distribution, does it meet government requirements?	10
Q#3 (a)	Describe the advantages and disadvantages of control charts for attributes.	15
(b)	A process produces rubber belts in lots of size 2500. Inspection records on the last 20 lots reveal the following data.	10

Lot Number	Number of Nonconforming Belts	Lot Number	Number of Nonconforming Belts
1	230	11	456
2	435	12	394
3	221	13	285
4	346	14	331
5	230	15	198
6	327	16	414
7	285	17	131
8	311	18	269
9	342	19	221
10	308	20	407

PTO

	i. Compute trial control limits for a fraction nonconforming control chart. ii. If you wanted to set up a control chart for controlling future production, how would you use these data to obtain the center line and control limits for the chart?	
Q#4(a)	Take a sampling plan with $n_1 = 50, c_1 = 2, n_2 = 100, c_2 = 6$ If the incoming lots have fraction nonconforming $p = 0.05$ then what is the probability of final acceptance? Calculate the probability of rejection on the first sample?	15
(b)	Define Average Outgoing Quality (AOQ). How do we interpret it?	10
Q#5(a)	i. Derive an item-by-item sequential Sampling Plan for which $p_1 = 0.01,\ \alpha = 0.05,\ p_2 = 0.10,\ \beta = 0.10$ ii. Draw the OC curve for this plan.	15
(b)	Why an organization should implement ISO? Also describe the internal and external objectives of this organization.	10
Q#6 (a)	In a plan, 10 items were tested for 500 hours with replacement and an acceptance number of 1. Construct an OC-curve showing probability of acceptance as a function of mean life.	15
(b)	An acceptance sampling plan for life testing requires that a sample of 19 items be tested with replacement for 1000 hours. If not more than 7 failures occur, the lot is accepted, otherwise it is rejected. Assume that the probability of failure is constant. Compute the mean life for which: 1. The producer's risk of a lot rejection is 0.05. 2. The consumer's risk of a lot acceptance is 0.10	10
Q#7	Write a short note on any **Five** of the following: i. Multiple Sampling Plan ii. Average Time to Signal (ATS) iii. The Bath-tub Curve iv. Rectifying Inspection v. CUSUM chart vi. Warning Limits vii. Multiple Sampling Plans	5 each

UNIVERSITY OF THE PUNJAB

Part-II **A/2016**
Examination:- M.A./M.Sc.

Roll No. ....

Subject: Statistics **TIME ALLOWED: 3 hrs.**
PAPER: VI (iii) [Operations Research] **MAX. MARKS: 100**

NOTE: Attempt any Four questions. Graph paper will be provided on demand.

Q.1 (a) Define OR and it importance in daily life. Explain your understanding with practical examples
(b) Discuss the phases of OR.
(c) A firm manufactures two types of products A and B and sells them at a profit of Rs. 2.00 on type A and Rs. 3.00 on type B. Each product is processed on two machines G and H. Type A requires one minute of processing time on G and two minutes on H; type B requires one minute on G and one minute on H. The machine G is available for not more than 6 hours 40 minutes while machine H is available for 10 hours during any working day. Formulate the problem as a linear programming model. 10+5+10

Q.2 (a) What is Linear Programming? How it is applicable in our daily life?
(b) How many forms of linear programming have you browsed? Illustrate their properties with examples.
(c) Explain the optimality and feasibility conditions of dual simplex methods. 10+10+5

Q.3 (a) Explain what is degeneracy? Discuss with the following LP model
Max $X_o = 3X_1 + 9X_2$ Subject to $X_1 + 4X_2 \le 8$: $X_1 + 2X_2 \le 4$; $X_1, X_2 \ge 0$
(b) Explain the role of slack, artificial and surplus variables in providing the solution of linear programming. 15+10

Q.4 (a) Find optimal solution by Dual Simplex Method. Max $X_o = 2X_1 - X_2 + X_3$
Subject to $2X_1 + 3X_2 - 5X_3 \ge 4$; $-X_1 + 9X_2 - X_3 \ge 3$; $4X_1 + 6X_2 + 3X_3 \le 8$;
$X_1, X_2, X_3 \ge 0$
(b) Make a comparison between Big M-technique and Phase-II method. 15+10

Q.5 (a) What is Transportation model? Explain its components.
(b) Consider the Transportation problem having the following cost table

		Destination				
		1	2	3	4	Supply
	1	10	0	20	11	15
Source	2	12	7	9	20	25
	3	0	14	16	18	5
	Demand	5	15	15	10	

Solve by (i) North-West Corner Rule (ii) Least Cost Method (iii) VAM 10+15

Q.6 (a) What is game? Explain the principle of minimax to solve a game.
(b) Gave a graphical representation of the following game. Determine the value of the game and the optimal mixed strategy for the player who has two strategies.

	B			
	-4	8	2	-10
A	6	-2	0	5

10+15

Q.7 (a)What are the three estimates used in the context of PERT? How are the expected duration of a project and its SD calculated?
(b) Draw the network corresponding to the following data

Activity	A	B	C	D	E	F	G	H	I
	1-2	1-3	1-4	2-3	2-4	3-5	3-6	4-5	5-6
Expected time (in days)	3	4	14	10	5	4	6	1	1

Compute the following
(i) EST and EFT of each activity (ii) Critical path and project duration
(iii) TS and ES for each activity 10+15

UNIVERSITY OF THE PUNJAB

Part-II A/2016
Examination:- M.A./M.Sc.

Roll No. ....

Subject: Statistics
PAPER: VI (iv) [Part-A-Survey and Research Methods]

TIME ALLOWED: 3 hrs.
MAX. MARKS: 50

NOTE: Attempt any FOUR questions. All questions carry equal marks.

Q.1. What is a difference between census and sample survey? Explain types of survey in detail. Also, explain the advantages and limitations of survey.

Q.2. Define scale and its types. Also, explain different kinds of rating scales by giving at least two examples each.

Q.3. What are the qualities of good data collection? Explain sources of primary and secondary data.

Q.4. Explain the principles of wording, stating how these are important in questionnaire designing?

Q.5. Explain validity and its types. Also explain different measures of reliability.

Q.6. a) Explain open ended and close-ended question. Their advantage and disadvantages.

b) What are the ethical and other related issues arising in a survey?

Q.7. Write note on any three of the following:

a) Importance of literature review
b) Sampling frame and its effect on survey
c) Pretesting of a questionnaire
d) Processing of survey data

UNIVERSITY OF THE PUNJAB

Part-II A/2016
Examination:- M.A./M.Sc.

Roll No. ....

Subject: Statistics **TIME ALLOWED: 3 hrs.**
PAPER: VII (i) [Time Series Analysis] **MAX. MARKS: 100**

NOTE: Attempt any FOUR questions.

Q.1.a) Describe the four components of time series. **(5)**

b) Given the following time series:

t	1	2	3	4	5	6	7
Yt	8	11	10	9	9	10	11

(10)

Calculate the sample autocorrelation and partial autocorrelation at lag 1 and 2.

c) Describe the invertibility condition for a moving average process. Check the invertibility of the following two MA(1) processes **(10)**

Model-1: $Y_t = Z_t + 0.25Z_{t-1}$

Model-2: $Y_t = Z_t + 4Z_{t-1}$

Discuss your findings.

Q.2.a) Define autoregressive process and state its model. **(5)**

b) Show that the autocorrelation function of $Y_t = Y_{t-1} - \frac{1}{2}Y_{t-2} + Z_t$ is given by **(12)**

$$\rho_k = \left(\frac{1}{\sqrt{2}}\right)^k \left(\cos\frac{\pi k}{4} + \frac{1}{3}\sin\frac{\pi k}{4}\right), \quad k = 0,1,2,3,\dots$$

where $\{Z_t\}$ is a purely random process having zero mean and finite variance.

c) Derive the recursive rule for autocorrelation function and thus the Yule-Walker equations for an AR(p) process. **(8)**

Q.3.a) Derive the mean, variance and autocorrelation function of a moving average process of finite order q i.e. an MA(q) process. Discuss the behavior of the autocorrelation function you obtained for MA(q) process **(10)**

b) Show that the AR(2) process given by **(5)**

$$Y_t = Y_{t-1} + cY_{t-2} + Z_t$$

is stationary provided $-1 < c < 0$.

c) Show that the infinite order MA process $\{Y_t\}$ defined by **(10)**

$$Y_t = Z_t + \theta\sum_{i=1}^{\infty} Z_{t-i}$$

is non-stationary. Also show that $W_t = Y_t - Y_{t-1}$ is stationary. Find the autocorrelation function of W_t.

Q.4.a) Show that the ACF after lag 2 of following MA(2) process cuts off to zero. **(12)**

$$Y_t = Z_t - 0.5\, Z_{t-1} + 0.2\, Z_{t-2}$$

where $\{Z_t\}$ is a purely random process having zero mean and finite variance.

b) If $Y_t = \mu + Z_t + \theta Z_{t-1}$, where μ is a constant, show that the ACF does not depend on μ. **(8)**

P.T.O.

c) If $U_t = \phi U_{t-1} + w_t$ and $V_t = \phi V_{t-1} + z_t$ are two independent AR(1) processes then show that $Y_t = U_t + V_t$ is an AR(1) process, where w_t and z_t are independent purely random processes. **(5)**

Q.5.a) Describe the iterative procedure of obtaining the least squares estimates of an ARMA(1,1) process. **(10)**

b) If $\{Y_t\}_{t=1}^{n}$ follows an AR(p) process $Y_t = \sum_{i=1}^{p} \phi_i Y_{t-i} + z_t$ where $\{z_t\}$ is independently and normally distributed process with zero mean and finite variance σ_z^2 then show that log-likelihood function is **(15)**

$$\ln L = const. - \frac{n}{2}\ln\sigma_z^2 + \frac{1}{2}\ln M_p - \frac{1}{2\sigma_z^2}\left(\sum_{t=p+1}^{n}\left(Y_t - \sum_{i=1}^{p}\phi_i Y_{t-i}\right)^2 + X^T M_p X\right)$$

where $(Y_1, Y_2, \cdots, Y_p) \sim N_p(0, \Sigma_p)$ and $M_p = \sigma_z^2 \Sigma_p^{-1}$. Also suggest procedure to find the approximate maximum likelihood estimates of AR parameters.

Q.6.a) Show that the minimum mean squared error forecast with origin at n and lead time l is given by **(15)**

$$Y_n(l) = \frac{1}{\psi(B)}\left[\frac{\psi(B)}{B^l}\right]_+ Y_n$$

where $\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \ldots$ and $\{\psi_j$: j=1,2,...} are the weights of moving average representation.

b) Show that for an ARIMA(1,1,1) process: $(1-\phi B)\nabla Y_t = (1+\theta B)Z_t$, the forecast at origin t with lead time l is given by **(10)**

$$Y_t(l) = Y_t + \frac{\phi(1-\phi^l)}{1-\phi}(Y_t - Y_{t-1}) + \frac{\theta(1-\phi^l)}{1-\phi}Z_t$$

Q.7.a) Show that the covariance between forecast errors for different lead times but with same forecast origin is given by **(10)**

$$cov\big(e_n(l), e_n(l+j)\big) = \sigma_z^2 \sum_{i=0}^{l-1} \psi_i \psi_{i+j}$$

for $j \geq 0$.

b) An AR(1) model is fitted to an observed time series $\{Y_t\}_{t=1}^{100}$. Given the following model **(15)**

$$Y_t = \mu(1-\phi) + \phi Y_{t-1} + z_t,$$

where $\hat{\mu}=65$, $\hat{\phi}=0.7$. Find the forecasts of $Y_{100+l}: l = 1,2$ and the 95% forecast limits when $\sigma_z^2 = 5$, $Y_{100} = 62.6$.

UNIVERSITY OF THE PUNJAB

Part-II A/2016
Examination:- M.A./M.Sc.

Roll No. ....

Subject: Statistics **TIME ALLOWED: 3 hrs.**
PAPER: VII (ii) [Multivariate Analysis] **MAX. MARKS: 100**

NOTE: Attempt any FOUR questions.

Q1.		Write short note on any **FIVE** of the following carrying equal marks: i. Factor Rotation ii. Spectral Decomposition of the Matrices iii. Sampling properties of Eigen Values and Eigen Vectors. iv. Discriminant Analysis v. Canonical Correlations Analysis vi. Principal Component Analysis vii. Multivariate Analysis viii. Positive Definite and Semi-positive Definite Matrices	*(25)*
Q2.	**(a)**	If $\mathbf{X}$ denotes a $(p\times 1)$ column vector of random variables, $\boldsymbol{\mu}$ is a column vector of constants and Σ is a positive definite matrix, then find the value of k such that $f(\mathbf{X}) = k\exp\left[-\frac{1}{2}(\mathbf{X}-\boldsymbol{\mu})'\,\Sigma^{-1}(\mathbf{X}-\boldsymbol{\mu})\right]$ is a pdf.	*(10)*
	(b)	The random vector $\mathbf{X}' = [X_1 \quad X_2 \quad X_3 \quad X_4]$ has a Multivariate Normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ given by: $\boldsymbol{\mu} = \begin{bmatrix} 4 \\ 0 \\ -3 \\ 7 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 7 & -1 & 0 & 3 \\ & 8 & -2 & 6 \\ & & 12 & 9 \\ & & & 3 \end{bmatrix}$ Suppose $\mathbf{Y}_1' = [X_1 \quad X_2]$ and $\mathbf{Y}_2' = [X_3 \quad X_4]$ are the subvectors of $\mathbf{X}$ then find (a) $E\left(\mathbf{Y}_1/\mathbf{Y}_2\right)$ (b) $Cov\left(\mathbf{Y}_1/\mathbf{Y}_2\right)$	*(15)*
Q3.	**(a)**	Fifty bars of soap are manufactured in each of two ways. Two characteristics X_1 = Lather and X_2 = Mildness are measured. The summary statistics for bars produced by method 1 and 2 are: $\bar{\mathbf{X}}_1 = \begin{bmatrix} 8.3 \\ 4.1 \end{bmatrix}, \bar{\mathbf{X}}_2 = \begin{bmatrix} 10.2 \\ 3.9 \end{bmatrix}, \mathbf{S}_1 = \begin{bmatrix} 2 & 1 \\ 1 & 6 \end{bmatrix}, \mathbf{S}_2 = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}$ (a) Test at 5% level that $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ (b) Find a 95% confidence region for $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ (c) Find 95% simultaneous confidence intervals for the differences in the mean components. (d) Find 95% Bonferroni Confidence Intervals for the Mean differences $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$.	*(18)*
	(b)	Let $\mathbf{W} \sim \mathbf{W}_p(f,\Sigma,\mathbf{M})$. If $\mathbf{C}$ is any $(p\times q)$ matrix of constants, then show that $\mathbf{C}'\mathbf{W}\mathbf{C} \sim \mathbf{W}_q(f,\mathbf{C}'\Sigma\mathbf{C},\mathbf{M}\mathbf{C})$.	*(07)*

P.T.O.

Q4 (a) The following is the covariance matrix for two random variables X_1 and X_2: *(13)*

$$\Sigma = \begin{bmatrix} 180.67 & 39.91 \\ & 67.31 \end{bmatrix}$$

Obtain the Principal Components Y_1 and Y_2. Also find the covariance matrix of principal components and prove that $\mathbf{B'B} = \mathbf{I}$.

(b) Suppose the random variable X_1, X_2, X_3 have the covariance matrix *(12)*

$$\begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

and the eigen-value, eigen-vector pairs are

$$\lambda_1 = 5.83, \qquad e_1' = [0.383 \quad -0.924 \quad 0]$$

$$\lambda_2 = 2.00, \qquad e_2' = [0 \quad 0 \quad 1].$$

$$\lambda_2 = 0.17, \qquad e_3' = [0.924 \quad 0.383 \quad 0]$$

(a) Find the Principal Components Y_1, Y_2, Y_3.
(b) Correlation of Y_1 with X_1 & X_2.
(c) Percentage of total variance explained by the first two Principal Components.

Q5. (a) Explain the method of factor analysis, indicating the assumptions involved. *(15)*

(b) Perform an appropriate factor analysis for the following matrix: *(10)*

$$\Sigma = \begin{bmatrix} 1.0 & 0.9 & 0.7 \\ & 1 & 0.4 \\ & & 1.0 \end{bmatrix}$$

Q6. (a) Suppose that population 1 $P_1 \sim N(\mu_1, \sigma_1^2)$ and population 2 $P_2 \sim N(\mu_2, \sigma_2^2)$. Discuss the maximum likelihood discriminant rule. *(10)*

(b) Consider $n_1 = n_2 = 50$ observations on two species of Iris gave the following mean vectors and covariance matrices: *(15)*

$$\bar{\mathbf{X}}_1 = \begin{bmatrix} 5.006 \\ 3.428 \end{bmatrix}, \bar{\mathbf{X}}_2 = \begin{bmatrix} 5.936 \\ 2.770 \end{bmatrix}, \mathbf{S}_1 = \begin{bmatrix} 0.1218 & 0.0977 \\ & 0.1408 \end{bmatrix}, \mathbf{S}_2 = \begin{bmatrix} 0.2611 & 0.0835 \\ & 0.0965 \end{bmatrix}$$

Find the Discriminat function and allocate the new observations $\bar{\mathbf{X}}' = [5.296 \quad 3.213]$ to any of these species.

Q7. (a) What is Canonical Correlation? Derive the canonical correlations and canonical variables. *(15)*

(b) The covariance matrix for four standardized variables Z_1, Z_2, Z_3, Z_4 is, *(10)*

$$\boldsymbol{\rho} = \begin{bmatrix} 1 & 0.4 & 0.5 & 0.6 \\ & 1 & 0.3 & 0.4 \\ & & 1 & 0.2 \\ & & & 1 \end{bmatrix}$$

Let $\mathbf{Z}_1' = [Z_1 \quad Z_2]$ and $\mathbf{Z}_2' = [Z_3 \quad Z_4]$. Find canonical correlation between $\mathbf{Z}_1'$ & $\mathbf{Z}_2'$.

Also find the first pair of canonical variates.