UNIVERSITY OF THE PUNJAB

M.A./M.Sc. Part – II Annual Exam – 2019

Roll No. ....

Subject: Statistics **Paper: I (Statistical Inference)** **Time: 3 Hrs.** **Marks: 100**

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

Q.1 a) Define unbiasedness, write four types of unbiasedness, and compare them. **(13)**

b) Let $Y_1<Y_2<Y_3$ be the order statistics of random sample of size 3 from uniform distribution over the interval $[0,\theta]$. Show that $4Y_1$, $2Y_2$ and $(4/3)Y_3$ are all unbiased estimators for θ with respective variances $3\theta^2/5, \theta^2/5$ and $\theta^2/15$. **(12)**

Q.2 a) State and prove Cramer-Rao's inequality for minimum variance, under which condition(s) it fails to give minimum variance bound; state the condition(s) concerned. **(12)**

b) Find the MVB estimators of the parameters of Normal distribution, also find their variances and co-variances. **(13)**

Q.3 a) Find the Maximum likelihood estimator and moment estimators of the parameters of log-normal probability distribution. **(12)**

b) If $X \sim N(\mu,\sigma^2)$ for a random sample of size n then find MLE of a point A, such that $\int_A^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx = 0.05$ also find the minimum variance unbiased estimator of A. **(13)**

Q.4 a) Let $x_1, x_2, \ldots\ldots x_n$ be a random sample from beta distribution of the form $f(x;\theta) = \theta(1+\theta)(1-x)x^{\theta-1}$, $0<x<1$ for some $\theta>0$, estimate θ by method of moments. Also find the variance of θ. **(10)**

b) Let there are 3n observations $x_1, x_2 \ldots x_n$, $y_1, y_2 \ldots y_n$ and $z_1, z_2 \ldots z_n$ with same unknown variance σ^2. The mean values of the observations are given by $E(x_i) = \theta_1 + 2\theta_2 + 3\theta_3$, $E(y_i) = 2\theta_1 + 3\theta_2 + \theta_3$, $E(z_i) = 3\theta_1 + \theta_2 + 2\theta_3$ i=1,2,3...n. where $\theta_1, \theta_2, \theta_3$ are unknown parameters. Apply the least square method to derive estimates of contrasts $(\theta_1 - \theta_2), (\theta_2 - \theta_3), (\theta_3 - \theta_1)$ by using $\theta_1 + \theta_2 + \theta_3 = 0$ and obtain the unbiased estimate of σ^2. Also prove that $V(\hat{\theta}_i) = 14\sigma^2/9n$. Find the variance of each contrast, what do you conclude? **(15)**

Q.5 a) Compare χ^2 (chi-square), minimum chi-square and modified chi-square methods in statistical inference; Under which situations they are applied? **(07)**

b) Based on a random sample of size 'n' from the density $f(x/\theta) = 1/\theta$ with prior distribution as $g(\theta) = 1$, $0<\theta<1$ obtain the Baye's estimator for θ with respect to the squared loss function $\ell(\theta,t) = (t-\theta)^2/\theta^2$. **(10)**

c) What is the difference between Bayesian and Classical inference? Explain. What is Bayes' estimator? Differentiate between prior and posterior density. **(08)**

Q.6 a) Prove that the approximate values of k_0 and k_1 in SPRT (Sequential probability ratio test) are $\alpha_a/(1-\beta_a)$ and $(1-\alpha_a)/\beta_a$. **(08)**

b) State and prove the Neyman-Pearson Lemma. **(12)**

c) What do you mean by 95% confidence interval? Define confidence region? **(05)**

Q.7 a) What do you mean by Pivotal Quantity? Explain with the help of any suitable example. **(08)**

b) Construct large sample confidence interval for unknown parameter of Poisson distribution. **(08)**

c) Write the short notes on **(09)**

i) Most power test ii) confidence belt iii) shortest confidence interval

UNIVERSITY OF THE PUNJAB

M.A./M.Sc. Part – II Annual Exam – 2019

Roll No. ....

Subject: Statistics **Paper: II (Regression Analysis and Econometrics)** **Time: 3 Hrs.** **Marks: 100**

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

Q.1. What is Econometrics? Discuss its types, methodology and functions.

Q.2.a) Consider the model $\underline{Y} = X\,\underline{\beta} + \underline{\in}$ *Such that* $E\underline{\in} = 0,\ E\underline{\in\in'} = \sigma^2 V$. Derive best linear unbiased estimator of $\underline{\beta}$.

b) Observations Y_i are related to fixed quantities X_i *and* U_i by the relation $Y_i = \alpha + \beta X_i + U_i$ *where* U_i are independent quantities from the same population with same variance. An estimate of β is $\frac{1}{25}(Y_1 - Y_2 - Y_3 - Y_4 + Y_5 + Y_6)$. If the values of X_i are 11, 21, 32, 44, 50 and 61, find sampling variance of this estimate and compare with it the sampling variance of the OLS estimator of β.

Q.3.a) What do you know about instrumental variable? Show that the instrumental variable estimate is consistent and find estimated asymptotic variance of instrumental variable estimate.

b) Using recursive formula for equally speed orthogonal polynomials or otherwise, devrive upto 3rd degree orthogonal polynomials.

Q.4.a) Discuss two stage least squares method and its features in simultaneous equations models.

b) Consider the model $Y_{1t} = \alpha_1 + \alpha_2 Y_{2t} + \alpha_3 X_{1t} + u_{1t}$; $Y_{2t} = \beta_1 + \beta_2 Y_{1t} + \beta_3 X_{2t} + u_{2t}$

(i) Identify the structural equations of the model.

(ii) Estimated reduced form equations are:

$\hat{Y}_{1t} = 4 + 3X_{1t} + 8X_{2t}$; $\hat{Y}_{2t} = 2 + 6X_{1t} + 10X_{2t}$

Obtain consistent Estimates of structural parameters.

Q.5.a) Define Autocorrelation. Discuss assumptions and procedure of Durbin-Watson test for autocorrelation.

b) What measures should be taken to avoid multicollinearity.

Q.6.a) What is Heteroskedasticity? Discuss (i) some reasons which causes heteroskedasticity, and (ii) various procedures / assumptions to remove heteroskedasticity.

b) Discuss Wald's, Bartlett's and Durbin's instrumental variables.

Q.7. Explain the following:

(i) Specification errors
(ii) Sampling distribution of error sum of squares
(iii) Stepwise regression
(iv) Simultanous equations bias
(v) The Almon approach to distributed-Lag models

UNIVERSITY OF THE PUNJAB

M.A./M.Sc. Part – II Annual Exam – 2019

Roll No. ....

Subject: Statistics **Time: 3 Hrs.** **Marks: 75**

Paper: III (Part-A) [Data Processing and Computer Programming]

NOTE: Attempt any FOUR questions.

Q.1 **(a)** Differentiate between source program and object program..

(b) What are the uses of following DOS commands, explain the following commands by giving examples

(i) dir **(ii)** md **(iii)** chkdsk **(iv)** format

(c) Write an algorithm and flowchart to print out the odd numbers between 1 to 300 and their squares. (6+8+5)

Q2 **(a)** In FORTRAN (FTN) language, what IF, THEN and DO WHILE statements? Give examples.

(b) Write a FTN program that generates and finds the solution of following series.

(i) $1-\frac{1}{2}+\frac{1}{3}-\ldots\ldots\ldots\ldots+\frac{1}{50}$

(ii) $\frac{2}{1}*\frac{4}{3}*\frac{6}{5}*\ldots\ldots\ldots\ldots*\frac{50}{49}$, where "*" represents multiplication sign.

(c) Write the following statements in FORTRAN

(i) $\text{Sin}(x)+\tan(x)$ **(ii)** $\sqrt{cd+|ab-4|}$

(iii) $\frac{1}{\sigma\sqrt{2\Pi}}e^{\frac{-(x-\mu)^2}{2\sigma}}+3.0\sqrt{(x/y)}$ **(iv)** $5+\frac{at}{bt+\frac{1}{e}}$

(6+8+5)

Q3 **(a)** Write a program in FORTRAN to calculate and prints the mean, variances, simple linear regression line and the correlation coefficient between two sets of data.

(b) i. Write a program in FORTRAN to read in an integer N > 4 and n >1, and to determine its Permutation.

ii. With reference to FORTRAN, explain BLOCK IF statement with example.

iii. Write a subroutine that sort the data in descending order.

(10+9)

Q4 **(a)** Write Subroutine subprograms which read all elements of two matrices A and B of any order; find the product of these two matrices and prints their product. Use these subroutines in a main program unit.

(b) Determine the output of the following programs:

```
C  First program
   I=4
   K=6
   L=K+2*I
   I=2*L+I/2
   K=K/4
   L=I+K+L
   WRITE(*,*) I,K,L
   STOP
   END
```

```
C  Second program
   A=6.4
   B=600.2
   W=21.3
   Z=A+B*W
   WRITE(*,*) A,B,Z
10 FORMAT(T10,F6.2,//,T10,F6.2,//T10,F6.2//T10,F8.2)
C
   STOP
   END
```

(c) There are N students in a class. The name and his/her scores in a semester examination consisting of four papers are given. Write a Fortran program to print the names of a student with respect to their GPA in descending order.

(6+6+7)

Q5 **(a)** With particular reference to C-Language,

i. Write the role of header files and pre-processor directives.

ii. Differentiate between Structure and Union, give examples.

iii. Differentiate between actual and formal arguments of a function.

iv. Differentiate between global and static variable.

v. What is the objective of header files before main() in C-language?

(b) With reference to the C- language, What are pointers? Write the logic that how they are used.

(15+4)

Q6 **(a)** Distinguish between a **switch** statement and an **if-else** statement in C language.

(b) Write a C- Program to find out the four raw moments of an array.

(c) Write a program in C to draw a checker's board on the screen.

(7+6+6)

Q7 **(a)** Write a recursive function to find binary equivalent of a number.

(b) Write a program in C which stores N values and then calls a function named **Array( )** to calculate the average, variance, minimum value and maximum value .

(c) Write a loop that will calculate the sum of every forth integer, beginning with i=3 (i.e. calculate the sum 3+7+11+...) for all values of i that are less than 500, using

i. while statement

ii. do-while statement

iii. for statement

(7+6+6)

UNIVERSITY OF THE PUNJAB

M.A./M.Sc. Part – II Annual Exam – 2019

Roll No. ....

Subject: Statistics **Paper: VI (iv) [Part-A-Survey and Report Writing]** **Time: 3 Hrs.** **Marks: 50**

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

Q.1. What is meant by sample survey? What are the advantages and limitations of survey? What steps should be taken into consideration for a successful survey?

Q.2. What do you understand by the term error? Explain different types of errors that affect the accuracy of the survey.

Q.3. Discuss various methods of random sampling and situations in which they are used.

Q.4. What are qualities of "Good Data"? Explain its types and sources to collect data.

Q.5. Explain the principles of wording, stating how these are important in questionnaire designing?

Q.6.a) Differentiate between open ended and close-ended questions. Also, explain their advantage and disadvantages.

b) Explain ethics of conducting survey.

Q.7. Define the term 'Reliability'. Also explain its various types.

UNIVERSITY OF THE PUNJAB

M.A./M.Sc. Part – II Annual Exam – 2019

Roll No. ....

Subject: Statistics **Paper: VI (i) [Statistical Quality Control]** **Time: 3 Hrs.** **Marks: 100**

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

Q#1 (a)	Describe the following: 1. Operating Characteristic (OC) Curve 2. Specification Limits	10
(b)	What is process capability ratio? How do we interpret the value of PCR?	15
Q#2	Control charts for $\bar{x}$ *and* R are to be established to control the tensile strength of a metal part. Assume that tensile strength is normally distributed. Thirty samples of size $n = 6$ parts are collected over a period of time with the following results: $$\sum_{i=1}^{30} \bar{x}_i = 6000 \text{ and } \sum_{i=1}^{30} R_i = 150$$ i. Calculate control limits for $\bar{x}$ *and* R control charts. ii. Both charts exhibit control. The specifications on tensile strength are 200 ± 5. what fraction non-conforming would result? iii. For the above $\bar{x}$ chart, find the β-risk when the true process mean is 199. iv. Assuming that if an item exceeds the upper specification limit it can be reworked, and if it is below the lower specification limit it must be scrapped, what percent scrap and rework is the process now producing? v. What are your conclusions regarding process capability?	25
Q#3 (a)	What to do if we want to deal with low defect levels?	10
(b)	As a part of an overall quality improvement program, a textile manufacturer decides to initiate a chart to monitor the number of imperfections found in each bolt of cloth inspected. The data, from last 25 inspections of the bolts of cloth, are recorded as 14, 5, 10, 19, 0, 6, 2, 9, 8, 7, 3, 12, 1, 22, 1, 6, 14, 8, 6, 9, 7, 1, 5, 12, 4 i. From these data, compute 3-sigma trial control limits. ii. Make control chart for number of imperfections. iii. What do you suggest about the process? iv. If manufacturer wishes to specify the trial control limits for the next period on the basis of current data available, suggest him the limits for the next period.	15
Q#4(a)	State the stipulated management principles for organizations considering an Environmental Management System under ISO 14000.	10
(b)	A single sampling plan uses a sample size of 15 items with an acceptance number 1. Using Hypergeometric distribution, draw the OC-curve for the lot having 50 articles.	15
Q#5 (a)	Take a sampling plan with $n_1 = 50, c_1 = 2, n_2 = 100, c_2 = 4$ If the incoming lots have fraction nonconforming $p = 0.05$ then what is the probability of acceptance on the first sample? What is the probability of final acceptance?	15

P.T.O.

(b)	Discuss the important considerations while forming lots for inspection in acceptance sampling procedures.	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.	10
(b)	Use the following data to set up short run $\bar{x}$ *and* R charts using the DNOM approach. The nominal dimensions for each part are $N_A = 50,\quad N_B = 25$	15

Sample No.	Part No.	M_1	M_2	M_3
1	A	50	51	52
2	A	49	50	51
3	A	48	49	52
4	A	49	53	51
5	B	24	27	26
6	B	25	27	24
7	B	27	26	23
8	B	25	24	23
9	B	24	25	25
10	B	26	24	25

Q#7	Write a short note on any **Five** of the following: i. Reliability and Life Testing ii. Average Outgoing Quality iii. Bath-tub Curve iv. Producer's and Consumer's Risk v. Average Run Lengths vi. Warning Limits	5 each

UNIVERSITY OF THE PUNJAB

M.A./M.Sc. Part – II Annual Exam – 2019

Roll No. ....

Subject: Statistics **Paper: VI (iii) [Operations Research]** **Time: 3 Hrs.** **Marks: 100**

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

Q.1. a) What are the situations where operations research techniques will be applicable?

b) Discuss the various phases in solving an O.R problem. **(12+13)**

Q.2. What do you mean by an optimal basic feasible solution to a Linear programming problem? Is the solution $x_1 = 1$, $x_2 = 1/2$, $x_3 = x_4 = x_5 = 0$, a basic solution of the equation $x_1 + 2x_2 + x_3 + x_4 = 2$ and $x_1 + 2x_2 + 1/2x_3 + x_5 = 2$ (25)

Q.3. Explain the transportation model and solve the transportation problem by using north-west corner rule and least cost method for which the cost, origin, availabilities and destination requirements are given below:

	D_1	D_2	D_3	D_4	D_5	D_6	a_i
O_1	1	2	1	4	5	2	30
O_2	3	3	2	1	4	3	50
O_3	4	2	5	9	6	2	75
O_4	3	1	7	3	4	6	20
b_j	20	40	30	10	50	25	

(25)

Q.4. What do you know about Network (Arrow Diagram) logic? Discuss it in detail with examples. (25)

Q.5. An aircraft company users rivets at an approximate customer rate of 2500 kg per year. The rivets cost Rs. 30 per kg and the company personnel estimate that it costs Rs.130 to place an order and the inventory carrying cost is 10% per year. How frequently should orders for rivets be placed and what quantities should be ordered? (25)

Q.6. a) Explain the following:

(i) Objective of games, (ii) Saddle point, (iii) Mixed strategy, (iv) Rectangular game.

b) Find the solution of the following game:

Player B

Player A			
	1	3	11
	8	5	2

(12+13)

Q.7. a) A new television set arrives for inspection every 5 min and is taken by a quality control engineer on a first come first served basis. There is only one engineer on duty and it takes exactly 6 min to inspect each new set. Determine the average number of sets waiting to inspected over the first 45 min of a shift, if there are no sets awaiting inspection at the beginning of the shift.

b) Explain the deterministic models of queuing theory. (13+12)

UNIVERSITY OF THE PUNJAB

M.A./M.Sc. Part – II Annual Exam – 2019

Roll No. ....

Subject: Statistics **Paper: VII (i) [Time Series Analysis]** **Time: 3 Hrs.** **Marks: 100**

NOTE: Attempt any FOUR questions.

Q.1 (a) Define the following:

i. Stochastic Process
ii. Stationarity
iii. Invertibility
iv. Portmanteau test (10)

(b) Consider the following time series of 10 observations:

t	1	2	3	4	5	6	7	8	9	10
Y_t	10	17	23	27	29	30	35	33	39	42

i. Plot the given time series.
ii. Calculate the mean and variance of Y_t and Y_{t-1} and discuss the stationarity of mean and variance.
iii. Calculate the autocorrelation at lag 1. (15)

Q.2 (a) Define the following stochastic processes:

i. Moving average process
ii. Seasonal ARIMA model (4)

(b) Find the mean, variance and autocorrelation function of the MA(2) process given by

$$Y_t = Z_t + 0.7\, Z_{t-1} - 0.2\, Z_{t-2}$$

where $\{Z_t\}$ is a purely random process having zero mean and finite variance.. (12)

(c) Show that for the ARMA(1,1) process $Y_t = \phi Y_{t-1} + Z_t + \theta Z_{t-1}$

$$\rho_k = \begin{cases} (1+\phi\theta)(\phi+\theta) \big/ (1+\theta^2+\phi\theta) & k=1 \\ \phi\rho_{k-1} & k=2,3,\cdots \end{cases}$$ (9)

Q.3 (a) Define the autoregressive process. Find the mean, variance autocorrelation function and partial autocorrelation function of an AR(1) process. Discuss the behavior of these autocorrelation functions. (12)

(b) Derive the stationarity conditions on the parameters for an AR(2) process. (8)

(c) Show that for an AR(p) process $Y_t = \sum_{i=1}^{p} \varphi_j Y_{t-j} + Z_t$,

$$\sigma_Y^2 = \frac{\sigma_Z^2}{1 - \sum_{i=1}^{p} \varphi_i \rho_i}$$

Where ρ_k is the autocorrelation at lag k. (5)

Q.4 (a) Describe the principle of parsimony. Mixed ARMA models are generally parsimonious then AR and MA models, discuss. (5)

(b) Consider the AR(2) process given by

$$Y_t = Y_{t-1} - \frac{1}{2} Y_{t-2} + Z_t,$$

where Z_t is a purely random process with zero mean and finite variance. Show that

$$\rho_k = \left(\sqrt{2}\right)^{-k}\left(\cos\frac{\pi k}{4} + \frac{1}{3}\sin\frac{\pi k}{4}\right); \qquad k = 0,1,2,\cdots$$ (12)

(c) For each of the following model, determine whether the model is stationary and/or invertible:

i. $Y_t = 0.3 Y_{t-1} + Z_t$
ii. $Y_t = 0.5 Y_{t-1} + Z_t - 1.3 Z_{t-1} + 0.4 Z_{t-2}$ (8)

P.T.O.

Q.5 (a) Describe how autocorrelation function and partial autocorrelation function of a sample time series are helpful in identifying the process which possibly has generated the observed time series. (6)

(b) Describe the iterative procedure of obtaining the least squares estimates of an MA(1) process with non-zero mean. Suggest some method for selection of appropriate starting values of least square estimate. (12)

(c) Derive the Yule-Walker estimates of an AR(2) process; $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + Z_t$. (7)

Q.6 (a) Given the following AR(2) process

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + Z_t$$

Derive the loss function for the Maximum likelihood estimation of AR parameters. Also obtain the MLEs (approximate) of ϕ_1, ϕ_2 and σ_Z^2. (15)

(b) An AR(1) model is fitted to an observed sample time series $\{Y_t\}_{t=1}^{50}$ of 50 observations and the following residuals autocorrelations are calculated.

k	1	2	3	4	5	6	7	8	9	10
r_k	0.25	0.18	0.05	-0.01	0.03	0.01	0.05	-0.04	0.01	0.02

Test the goodness of fit for AR(1) model using Ljung-Box portmanteau (Q_m^*) test with maximum lags in test $m = 5$. (10)

Q.7 (a) Show that for an AR(1) process $Y_t = \phi Y_{t-1} + z_t$, for an observed time series $\{Y_t\}_{t=1}^{n}$ the minimum mean squared error forecast of Y_{n+l} with origin at n and lead time l is given by

$$Y_n(l) = \phi^l Y_n.$$

Also show that

i. $Y_n(l)$ is an unbiased estimate of Y_{n+l}.

ii. $Var(e_n(l)) = \sigma_z^2\left(\dfrac{1-\phi^{2l}}{1-\phi^2}\right)$, where $e_n(l)$ is the forecast error. (15)

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

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

where $\hat{\mu} = 65$, $\hat{\phi} = 0.7$. Find the forecasts of $Y_{100+l} : l = 1,2$ and the 95% forecast limits when $\sigma_z^2 = 5$, $Y_{100} = 62.6$. (10)

UNIVERSITY OF THE PUNJAB

M.A./M.Sc. Part – II Annual Exam – 2019

Roll No. ....

Subject: Statistics **Paper: VII (ii) [Multivariate Analysis]** **Time: 3 Hrs.** **Marks: 100**

NOTE: Attempt any FOUR questions.

Q1. For the data matrix,

$$X = \begin{bmatrix} 1 & 2 & 5 \\ 4 & 1 & 6 \\ 4 & 0 & 4 \end{bmatrix}$$

Calculate the sample mean, variances and covariances of the linear combinations $b^T X$ and $c^T X$ where

$$b^T X = [2 \;\; 2 \;\; -1][X_1 \;\; X_2 \;\; X_3]^T,$$

$$c^T X = [1 \;\; -1 \;\; 3][X_1 \;\; X_2 \;\; X_3]^T.$$

(25)

Q2. If a random sample of size n is taken from a multivariate normal population of random vector $\mathbf{x}$ of order $(p \times 1)$ with mean $\boldsymbol{\mu}$ and covariance matrix Σ. Find the maximum likelihood estimators of $\boldsymbol{\mu}$ and Σ. (25)

Q3. The sample mean and covariance matrix of a sample of 42 observations drawn from a multivariate normal distribution are,

$$\bar{X} = \begin{bmatrix} 0.564 \\ 0.603 \end{bmatrix}, \quad S = \begin{bmatrix} 0.0144 & 0.0117 \\ 0.0117 & 0.0146 \end{bmatrix}$$

(i) Obtain 95% confidence ellipsoid for $\boldsymbol{\mu}$ and use this to test the hypothesis $\boldsymbol{\mu}_o^T = [9, 5]$.

(ii) Calculate 95% simultaneous T^2 confidence intervals for μ_1 and μ_2.

(iii) Calculate 95% Bonferroni simultaneous confidence intervals for μ_1 and μ_2 and compare them with the results obtained in (ii).

(10+6+9)

Q4.

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

Carryout principal component analysis for Σ. Calculate percentage of explained variation by each principal component. (25)

Q5.

(i) Let $X = [X_1, X_2, ..., X_p]$ have covariance matrix Σ with eigen value and eigen vector pairs (λ_i, e_i) for $i = 1, 2, ..., p$. Let $Y_i = e_i^T x$ be the ith principal component. Show that $\sum_{i=1}^{p} Var(X_i) = \sum_{i=1}^{p} Var(Y_i)$.

(ii) The eigen values and eigen vectors of a correlation matrix are

$\lambda_1 = 1.73$, $e_1 = [-0.11, 0.58, 0.23, -0.47, -0.33, -0.52]^T$
$\lambda_2 = 1.34$, $e_2 = [0.52, 0.07, -0.68, -0.46, 0.23, -0.06]^T$
$\lambda_3 = 0.96$, $e_3 = [0.52, -0.36, 0.08, 0.02, -0.77, 0.01]^T$
$\lambda_4 = 0.86$, $e_4 = [-0.57, -0.11, -0.32, -0.43, -0.37, 0.48]^T$
$\lambda_5 = 0.67$, $e_5 = [-0.36, -0.47, -0.37, 0.13, -0.06, -0.70]^T$
$\lambda_6 = 0.44$, $e_6 = [0.04, -0.55, 0.49, -0.60, 0.31, -0.05]^T$

Assuming two-factor model, calculate loadings and specific variances using the principal component solution method. What proportion of total variation is explained by the first two common factors?

(8+17)

Q6. The following mean vector and covariance matrices are based on $n_1 = n_2 = 100$ observations drawn from multivariate normal populations with equal covariance matrices.

$$\bar{X}_1 = \begin{bmatrix} 6.213 \\ 3.133 \end{bmatrix}, \quad S_1 = \begin{bmatrix} 1.813 & 0.321 \\ 0.321 & 0.937 \end{bmatrix}$$

$$\bar{X}_2 = \begin{bmatrix} 7.412 \\ 5.321 \end{bmatrix}, \quad S_2 = \begin{bmatrix} 2.193 & 1.654 \\ 1.654 & 3.789 \end{bmatrix}$$

Allocate the new observation $x_0^T = [7.2, 3.1]$ to one of these populations. (25)

Q7. Derive the canonical correlations and canonical variables. (25)