R.B.I.S.B. (Officer in Grade- 'B' – DSIM) PY-2016

PAPER II – DESCRIPTIVE TYPE ON STATISTICS

(Maximum Marks – 100) (Duration – 3 Hours)

Instructions :

(1) The question paper consists of **six Sections**. The candidate may attempt any ***five*** questions selecting **not more than two** from any section. In case the candidate answers more than five questions, only the first five questions in the chronological order of question numbers answered will be evaluated and the rest of the answers ignored.

(2) **QUESTIONS FROM EACH SECTION SHOULD BE ANSWERED ON SEPARATE ANSWER-SCRIPT/SUPPLEMENTS. In other words, a candidate may require/use minimum 2 to 4 supplements, in addition to Answer script.**

(3) Supplement should be attached to the answer script, before returning.

(4) Each question carries 20 marks.

(5) Answers must be written either in *English* or in *Hindi*. However, all the questions should be answered in one language only. Answer-books written partly in *English* and partly in *Hindi* will not be evaluated.

(6) Each question should be answered on new page and the question number must be written on the top in left margin.

(7) The answers of parts of the same question, if any, should be written together. In other words, the answer of another questions should not be written in-between the parts of a question.

(8) The Name, Roll No. and other entries should be written on the answer-scripts at the specified places only and these **should not be written anywhere else on the answer script and supplements**.

(9) Candidate should use only **Blue** or **Black** ink ball point pen to write the answers.

(10) No reference books, Text books, Mathematical tables, Engineering tables, other instruments or communication devices (including cellphones) will be supplied or allowed to be used or even allowed to be kept with the candidates. Violation of this rule may lead to penalties.

(11) **Use of non-programmable electronic calculator is permitted.**

(12) ALL ROUGH WORK MUST BE DONE IN THE LAST THREE OR FOUR PAGES OF THE ANSWER SCRIPT.

(13) Answers will be evaluated on the basis of logic, brevity and clarity in exposition.

(14) Marks will be deducted for illegible hand-writing.

[*Turn over*

PAPER II – DESCRIPTIVE TYPE ON STATISTICS

Section A: Probability and Sampling

1. (a) Consider a two-stage sampling with equal number of units in the first stage. Obtain the variance of an unbiased estimator of population mean when sample is selected without replacement in first-stage as well as in second-stage.

   (b) A population consists of 40 first stage units, each containing 12 second stage units. A random sample of 5 first stage units is selected and 3 second stage units are selected at random from each first stage units . The value of the variable under study were obtained as follows : —

First stage units	y		
1	16	31	21
2	25	27	27
3	17	22	21
4	32	29	24
5	18	16	26

   Estimate the population mean of y and obtain an estimate of its variance.

2. (a) Define ratio estimator of population total in case of simple random sampling without replacement and obtain its exact bias.

   (b) The following table shows the area under paddy (Y) and cultivated area (X) for the villages in the zone. A simple random sample of size 4 without replacement gives villages numbered 3, 5, 8 and 9 is selected. Obtain ratio estimate of total area under paddy along with its variance : –

Village number	1	2	3	4	5	6	7	8	9	10
X	1012	1181	780	815	1120	659	897	783	689	1217
Y	340	416	247	306	403	271	357	295	218	398

3. (a) State and prove Markov's inequality. Explain how Chebychev's inequality is a special case of Markov's inequality.

   (b) Let X be a uniformly distributed continuous random variable over the interval (0,1). Obtain lower bound for $P\left[\left|X-\frac{1}{2}\right| < \sqrt{\frac{1}{3}}\right]$.

Section B : Linear Models and Economic Statistics

4. (*a*) State and prove Gauss-Markov theorem under the model $\underset{n\times 1}{y} = \underset{n\times p}{X}\underset{p\times 1}{\beta} + \underset{n\times 1}{e}$ where $E(\underline{e}) = \underline{0}$ and $V(\underline{e}) = \sigma^2 I$, $p \leq n$ and $R(X) = p$

   (*b*) Samples of sizes n_1 and n_2 are drawn from two populations having means μ_1 and μ_2 respectively with common variance σ^2. Find BLUE of $l_1\mu_1 + l_2\mu_2$ along with its variance. Test the hypothesis $H_0: \mu_1 = \mu_2$ and give ANOVA.

5. (*a*) Write multiple linear regression model that has p predictor variables and n observations. Explain all the terms involved in the model. State the assumptions required and explain consequences of violation of any one of the assumptions with justification.

   (*b*) Consider multiple linear regression model with four predictor variables. Compute F statistic to test the hypothesis that predictor variables have no explanatory power. You are given that regression s.s.=23665352, residual s.s. (with 88 d.f.) = 22657938. Also compute R^2 and estimate of population variance.

6. (*a*) What do you understand by Index numbers? Discuss difficulties that arise in the construction of cost of living index numbers.

   (*b*) Compute Laspeyre's, Paasche's and Fisher's price index numbers for the following data using 1990 as base period. Prices are quoted in Rs. per Kg. and production is in quintals : —

Commodities	1990		1995	
	Price	Production	Price	Production
A	15	500	20	600
B	18	590	23	640
C	22	450	24	500

Section C: Statistical Inference

7. (*a*) Let T_0 be the uniformly minimum variance unbiased estimator (UMVUE) of $g(\theta)$ and v_0 be the unbiased estimators of zero. Assume that the second moment exist for all unbiased estimators of $g(\theta)$. Then prove that T_0 is UMVUE if and only if $Ev_0T_0 = 0 \ \forall\ \theta \in \Theta$.

(*b*) Let $X_1, X_2, \ldots, X_n$ be rvs with pdf

$$f(x|\theta) = \begin{cases} \frac{1}{2\theta} & ; \quad -\theta < x < \theta \\ 0 & ; \quad otherwise \end{cases}$$

Find UMVUE of $(i)\ \frac{\theta}{1+\theta}$ $\quad (ii)\ \frac{e^{\theta}}{\theta}$

8. (*a*) Define maximum likelihood estimator :—

(i) Let T be a sufficient statistics for the family of pdf (pmf) $f(x|\theta), \theta \in \Theta$. Prove that if an mle of θ exists and unique then it is a function of T.

(ii) Let the rv X have the hypergeometric distribution :—

$$P(x|N) = \begin{cases} \dfrac{\binom{M}{n}\binom{N-M}{n-n}}{\binom{N}{n}} & ; \quad Max(0, n-N+M) \le x \le Min\ (n, M) \\ 0 & ; \quad otherwise \end{cases}$$

Find the mle of N when M and n are known.

(*b*) Let X_1 and X_2 be independent exponential random variables with means λ_1 and λ_2 respectively. Let $Z_1 = \text{Min}\ (X_1, X_2)$ and

$$Z_2 = \begin{cases} 0 & ; \quad Z_1 = X_1 \\ 1 & ; \quad Z_1 = X_2 \end{cases}$$

Find mle of λ_1 and λ_2 in a sample size n based on $Z_1$and Z_2.

9. (*a*) Let $X_1, X_2, \ldots, X_n$ be iid rvs from $U(-\theta, \theta), \theta > 0, \theta \in \Theta$. Find UMP test for testing

(*iv*) $H_0: \theta = \theta_0$ against $H_1: \theta > \theta_0$

(*v*) $H_0: \theta = \theta_0$ against $H_1: \theta < \theta_0$

(*vi*) $H_0: \theta = \theta_0$ against $H_1: \theta \neq \theta_0$

(*b*) Ten science teachers have been ranked by a judge according to their teaching ability. All have taken national teacher's examination. The data are given as below :—

Teachers	1	2	3	4	5	6	7	8	9	10
Judge's Rank	7	9	4	2	6	8	3	1	5	10
Examination score	44	72	69	70	55	93	82	67	80	98

Calculate Spearman's rank correlation coefficient for judge's ranking and examination score. Can you conclude that there is a negative association at 5% level of significance. Also compute Kendall's tau. [Under Ho, P $(R > 0.552) = 0.05$]

Section D: Stochastic Processes

10. (*a*) Prove that a state E_j of a Markov Chain with t.p.m. $P = (p_{ij})$ is transient if and only if.

$$\sum_{n=1}^{\infty} p_{jj}^{(n)} < \infty$$

(*b*) Classify the states of the Markov Chain with the following transition probability matrix P.

$$P = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

11. (*a*) Define.— (*i*) irreducible (*ii*) aperiodic Markov Chain. State and prove ergodic theorem for a finite irreducible, aperiodic Markov Chain with t.p.m. $P = (p_{ij})$.

(*b*) Consider the three-state Markov Chain having the t.p.m. given below : —

$$\begin{bmatrix} 0 & 2/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}$$

Is the chain (*i*) irreducible (*ii*) aperiodic ? Find the stationary distribution for the above chain.

12. (*a*) Define .— (*i*) ρ_k: autocorrelation function at lag k.
(*ii*) ϕ_k:partial autocorrelation function at lag k.

For an autoregressive process of order p, derive expression for ϕ_k. Hence express ϕ_2 in terms of ρ_1 and ρ_2.

(*b*) From the data given below : —

k	1	2	3	4	5	6
r_k	0.84	0.73	0.61	0.54	0.47	0.46

Calculate the estimated partial autocorrelation functions $\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3$

SectionE : Multivariate Analysis

13. (*a*) Define canonical correlations and canonical variates. Give the procedure to obtain first canonical correlation and canonical variate.

(*b*) Let $\rho_{11} = \rho_{22} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$and $\rho_{12} = \begin{pmatrix} \rho & \rho \\ \rho & \rho \end{pmatrix}$, $-\frac{1}{3} < \rho < 1$ corresponding to the equal correlation structure where $X^{(1)}$ and $X^{(2)}$ each have two components . Determine the canonical variates corresponding to the nonzero canonical correlation.

14. (*a*) Test the equality of the mean vectors of two p-variate normal populations with same but unknown vaiance-covariance matrix Σ using Hotelling's T^2 statistic.

(*b*) Suppose that $n_1 = 30$ and $n_2 = 22$ observations are made on two random vectors X_1 and X_2 which are assumed to have bivariate normal distribution with a common covariance matrix Σ but possibly different mean vectors $\mu_1$and μ_2. The sample mean vectors and inverse of pooled variance matrix S_p are

$$\bar{X}_1 = \begin{pmatrix} -0.0065 \\ -0.039 \end{pmatrix}, \quad \bar{X}_2 = \begin{pmatrix} -0.2483 \\ -0.0262 \end{pmatrix} \quad S_p^{-1} = \begin{pmatrix} 131.158 & -90.423 \\ -90.423 & 108.147 \end{pmatrix}$$

Test the hypothesis $H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 \neq \mu_2$ at 5% level of significance. (Given $F_{tab}(2,49) = 3.19$ for $\alpha = 0.05$).

15. (*a*) Define population principal components. Show that $V(Y_1) \geq V(Y_2) \geq \cdots \geq V(Y_p)$ where Y_i denotes the i[th] principal component of p-variate population.

(*b*) Obtain the first two principal components Y_1 and Y_2 of the following correlation matrix :—

$$P = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}, \quad -\frac{1}{2} < \rho < 1$$

Obtain the proportion of variance explained by Y_1.

Section F : Numerical Analysis and Basic Computer Techniques.

16. (*a*) Find the approximate value of $I= \int_0^1 \frac{dx}{1+x}$, by using Trapezoidal rule. Obtain a bound for the errors. The exact value of I = ln 2= 0.693147 correct to six decimal points. Take h=1.

(*b*) Find the value of $I = \int_{-1}^{1} e^{-x^2} \cos x \, dx$, using Gauss-Legendre one point and two point formula.

17. (*a*) Obtain the least squares polynomial approximation of degree one and two for $f(x)=x^{½}$ on [0,1].

(*b*) Obtain the cubic spline approximation for the function defined by the data :—

x	0	1	2	3
f(x)	1	2	33	244

With M(0) =0, M(3)=0. Hence find an estimate of f(2.5).

18. (*a*) Answer the following questions :—
(*i*) What do you understand the looping concept in programming? Illustrate with examples.
(*ii*) What do you understand by distributed processing? Illustrate with example.

(*b*) Answer the following questions :—
(*i*) Describe any one searching algorithm.
(*ii*) Explain the best and worst case time complexities of the Bubble sort algorithm. Illustrate with examples.