UNIVERSITY OF THE PUNJAB

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

Roll No. .....

Subject: Statistics
PAPER: I (Statistical Methods)

TIME ALLOWED: 3 hrs.
MAX. MARKS: 100

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

Q.1.

a) Define exponential distribution. Name some of exponential distributions

b) A restaurant serves eight entries of fish, 12 of beef and 10 of poultry .if customer selects these entries randomly, what is the probability that two of the next four customers order fish entrees?

c) The painted light bulbs produced by a company are 30% red, 30% blue and 40% green. In a sample of 7 bulbs, find the probability that 2 are red, 3 is green and 2 are blue.

(7+7+6)

Q.2.

a) Write down the general procedure for testing of hypothesis.
b) Construct a confidence interval for variance of a normal population.
c) The following are the weights in ounces of 10 cans of peaches distributed by a certain company:

16.4, 16.1, 15.8, 17.0, 16.1, 15.9, 15.8, 16.9, 15.2 and 16.0.

i) Test the Null Hypothesis $\sigma^2 = 16$ against the alternative hypothesis that $\sigma^2 \neq 16$ at 5% level of significance.
ii) Find a 99% Confidence Interval for the variance of all such cans of peaches, distributed by this company and interpret.

(7+8+5)

Q.3.

a) Define the following
i) Type I and type II errors
ii) P-value
iii) Central limit theorem

b) Explain the relationship between Z, F and t distribution.

c) In an experiment to decide whether butter and margarine can be distinguished, 90 individuals were selected and each of them given three sandwiches, two containing butter and other one containing margarine. They were then asked which one of the three contain margarine, 38 correct identifications were made. Test the significance of the experiment.

(9+5+6)

Q.4.

a) Write a note on Chebyshev's theorem.

b) Given that $H_0: \mu \geq 200$ against the alternative $H_1: \mu < 200$, $n=100$, $\alpha = 0.023$ and $\sigma = 25$

i. Using the given information find the values of the sample mean that would lead to the acceptance of H_0.
ii. Compute β if μ is actually 191.
iii. Compute the power of test and interpret the result.

c) A random sample of 200 voters is selected and 120 are found to support an annexation suit. Find the 96% confidence interval for the fraction of the voting population favored the suit.

(5+5+10)

Q.5.

a) Explain the general procedure for Fisher's exact test.

b) Find the unbiased estimates of the least square regression line .Also find their covariance.

c) The back fat thickness (y) and slaughter weight (x) for certain kind of animals fed on two different rations are given below

The data for treatment I are:

Y_1	42	38	53	34	35	31	45	43
X_1	206	261	279	221	216	198	277	250

P.T.O.

The data for treatment II are

Y_2	33	34	38	33	26	28	37	31
X_2	167	192	204	197	181	178	236	204

Compute the two regression coefficients and test the hypothesis that they are estimates of common β.

. (6+6+8)

Q.6. a) What is non-parametric problem? Discuss the advantages and disadvantages of non-parametric tests.

b) In the manufacture of automobile gears, the following data were obtained for the daily number of defectives for a production of 100 parts per day:

22	23	21	17	38	25	20	16	14	26	15	25	29
26	25	26	32	26	23	30	31	43	18	26	21	22
18	24	16	16	17	16	28	32	18	26	36	42	36
34	24											

Is the number of runs above and below the median at $\alpha = 0.05$ regarded as a random sample?

c) Test the hypothesis by the KOLMOGOROV- SMIRNOV method, that the following sample: 0.36, 0.92, -0.56, 1.86, 1.74, 0.56, -0.95, 0.24, -0.15, -0.74, 0.32, 0.82, 0.70, -0.10, -1.06, 0.15, 0.55, -0.48, -0.49 come from a normal population with mean 0.5 and variance 1. Use $\alpha = 0.05$.

(7+4+4+5)

Q.7. a) Explain fully the following Non-parametric tests

(i) KOLMOGOROV- SMIRNOV method
(ii) The Mann- Whitney U test.
(iii) The Run test.
(iv) The Sign test

b) The nicotine content of two brands of cigarettes, measured in milligrams was found to be as follows:

Brand A	2.1, 4.0, 6.3 , 5.4 , 4.8 , 3.7 , 6.1 , 3.3
Brand B	4.1 , 0.6 , 3.1 , 2.5 , 4.0 , 6.2 , 1.6 , 2.2 , 1.9 , 5.4

Test the hypothesis at 5% level of significance, that the average nicotine contents of the two brands are equal against the alternative that they are unequal.

(12+8)

Q.8. **a)** What is meant by sequential analysis? Explain the general testing procedure of sequential test.

b) A random sample of 15 adults living in a small town is selected to estimate the proportion of voters favoring a certain candidate for mayor. Each individual was also asked if he or she was a college graduate. By letting Y and N designate the responses of "yes" and "no" to the education question, the following sequence was obtained

NNNNYYNYYNYNNNN

Use the runs test at the 0.05 level of significance to determine if the sequence supports the contention that the sample was selected at random.

(8+12)

Q.9.

a) Express f(a+nh) in terms of f(a) and successive differences of f(a) where n is an integer.

b) Interpolate the value of f(8) from the following data.

X	6	9	10	12	13	15
f (X)	13	18	21	30	32	35

By using
Lagrange's formula for unequal interval
Newton's formula for unequal interval

(6+ 14)

UNIVERSITY OF THE PUNJAB

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

Roll No. .....

Subject: Statistics **TIME ALLOWED: 3 hrs.**
PAPER: II (Probability and Probability Distributions) **MAX. MARKS: 100**

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

Q.1.a) A five sided dice with sides number {1,2,3,4,5} is constructed so that 1 and 5 occur twice as often 2 and 4 which occur 3 times as often as number 3. What is the probability that a perfect square occur when this dice is rolled. **(8)**

b) In the senior year of a high school graduating class of 100 students, 42 studied mathematics, 68 studied psychology, 54 studied history 22 studied both mathematics and history, 25 studied both mathematics and psychology, 7 studied history but neither mathematics nor psychology 10 studied all three subjects, and 8 did not take any of the three. If a student is selected at random, find the probability that **(8)**
i) Person enrolled in psychology takes all three subjects.
ii) A person not taking psychology is taking both history and mathematics.

c) A father, mother, 2 boys, and 3 girls are asked to line up for a photograph. Determine the number of ways they can line up if **(9)**
i) there are no restrictions
ii) the parents stand together
iii) the parents do not stand together

Q.2.a) Bowl A contains 2 red chips; bowl B contains two white chips; and bowl C contains 1 red chip and 1 white chip. A bowl is selected at random, and one chip is taken at random from that bowl. What is the probability of selecting a white chip? **(10)**

b) State and Prove Bayes Theorem. **(8)**

c) Find the c.d.f. of the random variable from the p.d.f. **(7)**

$$f(x)=\begin{cases} e^{-x} & 0<x<\infty \\ 0 & \text{otherwise} \end{cases}$$

and use it to evaluate $P(0.5 \le X \le 1)$.

Q.3.a) Find m.g.f. of Poisson distribution and generate first four mean moments from it. **(15)**

b) If X has a Geometric distribution then show that P(X>a+b / X>a)= P(X>b). **(10)**

P. T. O.

Q.4.

a) Obtain the moment generating function of Normal distribution. **(15)**

b) State and prove Law of large numbers. **(10)**

Q.5.

a) If (x, y) has a bivariate normal distribution then show that the marginal distribution of X is univariate normal distribution i.e. $X \sim N(\mu_x, \sigma_x^2)$. **(15)**

b) If $f(x,y) = x^2 + \frac{xy}{3}$ $\quad 0 \le x \le 1, \quad 0 \le y \le 2$

$= 0$ $\quad$ *elsewhere*

Find the value of ρ_{xy}. **(10)**

Q.6.a) Given that $X \sim N(\mu, \sigma^2)$. Obtain the distribution of $Y = |X|$ and also show that the area under the probability curve is unity. **(10)**

b) Derive F- distribution. **(10)**

c) If the random variable X has normal distribution with mean 'm' and variance 'σ^2'. Obtain the distribution of Y=aX+b. **(5)**

Q.7.a) If X and Y are random variables with p.d.f **(15)**

$f(x,y) = 4xye^{-(x^2+y^2)}\, 0 \le x; y \le \infty$

Find the p.d.f of Z where $Z = \sqrt{X^2 + Y^2}$.

b) Let $Y_1 < Y_2 < Y_3 < Y_4 < Y_5$ be the order Statistics of a random sample of size 5 from the distribution having p.d.f. **(10)**

$f(x) = e^{-x}, \quad 0 < x < \infty$

Show that the statistics $Z_1 = Y_2$ and $Z_2 = Y_1 - Y_2$ are statistically independent.

UNIVERSITY OF THE PUNJAB

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

Roll No. .....

Subject: Statistics **TIME ALLOWED: 3 hrs.**
PAPER: III (Design and Analysis of Experiments) **MAX. MARKS: 100**

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

Q.1 a) What are the basic principles of experimental design? 10

b) What is a contrast? Discuss the Test of Orthogonal Contrasts? 10

c) Define the relative efficiency. How would you calculate the relative efficiency of LS design relative to RCB design with rows as blocks? 5

Q.2 a) Prove that for CR design E[MSE]=σ^2 for random effect model (equal size) . 10

b) Given below the ANOVA table for RCB design, complete the table and also interpret the result. 5

S.O.V	d.f	SS	MS	F-ratio
B/W consignments	5	-	-	-
B/W Observers	-	13.13	-	-
Error	15	-	0.87	-
Total	23			

c) Given a $(P \times P)$ Latin Square design. Show that the estimate of one missing observation is 10

$$r = \frac{p(R+C+T)-2G}{(p-1)(p-2)}$$

Q.3 a) Given below is the sum of cross products for a RCBD with one covariate. Complete the analysis and draw conclusions. 13

S.O.V	d.f	XY	YY	XX
Blocks	2	1.9111	1.2667	0.9330
Treatments	14	10.3111	19.4000	68.1333
Error	28	10.7556	7.4000	13.7333

i. Is the regression of Y on X significance at 5% level?
ii. Construct ANOVA and write the inference.

b) What is a Latin Square Design? Construct a Latin Square Design for studying the effects of five factors. Outline the ANOVA table. 12

Q.4 a) In which condition we use a single replicate in a 2^k factorial design? And in this situation how can we estimate the error SS? 10

b) A soft drink bottler is interested in obtaining more uniform fill height in the bottles produced by his manufacturing process. The process engineer can control percentage carbonation, the operating pressure in the filler and the bottles produced per minute or the line speed during filling process. Two levels of carbonation, two levels of pressure and two levels of line speed are selected to study the effect of these factors in filling process. A factorial experiment with two replicates is conducted, the order in which the 16 observations are taken is determined randomly. The coded data are given below. Analysis the data using Yates technique for computing contrasts 15

P.T.O.

Percent Carbonation	Operating Pressure			
	25 psi		30 psi	
	Line Speed		Line Speed	
	200	250	200	250
10	-3 -1	-1 0	-1 0	1 1
12	0 1	2 1	2 3	6 5

Q.5 a) What do you mean by confounding? Explain the difference between complete and partial confounding with examples. 12

b) Complete ANOVA table for the following factorial experiment ABC is completely confounded. 13

Rep.I		Rep.II	
1	2	1	2
b=4	(1)=10	a=6	bc=6
a=6	ab=7	abc=9	ac=8
c=9	bc=8	c=8	(1)=11
abc=5	ac=6	b=7	ab=8

Q.6 a) Compare a split-plot experiment with factorial experiment. 8

b) Write the advantages and disadvantages of RCB design and split plot design. 7

b) In an experiment on rice, there are 4 methods of irrigation to be compared, on main plot, with 3 fertilizer mixtures on subplots. Two complete blocks were used. The Total SS, Block SS and Whole Plot Error are 3.0200, 0.0004 and 0.6509 respectively. The following table shows the sum of two blocks. Complete the ANOVA Table and draw your conclusion. 10

	I_1	I_2	I_3	I_4
F_1	4.68	4.34	3.78	4.67
F_2	5.02	4.91	4.01	4.67
F_3	6.00	5.16	4.10	5.45

i. Find the standard Error of the difference between two Fertilizer means.

ii. Find the standard Error of the difference between two Irrigation means.

Q.7 a) Explain incomplete block designs. Compare BIBD with PBIBD, giving examples. 12

b) Data on screen colour difference on a television tube measured in degrees Kelvin are to be compared for four operators. On a given day only three operators can be used in the experiment. A balanced incomplete block design gave results as follows: 13

Operators	Days			
	Monday	Tuesday	Wednesday	Thursday
1	173	174	-	171
2	-	175	167	172
3	183	175	168	-
4	175	-	172	175

UNIVERSITY OF THE PUNJAB

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

Roll No. .....

Subject: Statistics TIME ALLOWED: 3 hrs.
PAPER: IV (Sampling Techniques) MAX. MARKS: 100

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

Q#1 (a)	Discuss the role of sampling theory and various steps involved during the planning and execution of a survey.	12
(b)	Show that the variance of the estimate of population total $\hat{y} = N\bar{y}$ from a simple random sample is $V(\hat{y}) = N^2(1-f)\frac{S^2}{n}$.	08
Q#2(a)	Explain the idea of Inverse sampling.	10
(b)	If y_i, x_i are a pair of variates defined on every unit in the population and $\bar{y}, \bar{x}$ are the corresponding means from a simple random sample of size n, then prove that $$Cov(\bar{x},\bar{y}) = \frac{N-n}{nN}\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{X})(y_i-\bar{Y})$$	10
Q#3 (a)	Describe the different ways of allocation of sample size in strata.	10
(b)	If the terms in $\frac{1}{N_h}$ are ignored relative to unity, show that for estimated mean from stratified random sample of size n_h, $$V_{opt} \le V_{prop} \le V_{ran}$$ Where the optimum allocation is for fixed 'n'.	10
Q#4(a)	Define systematic sampling. In how many ways you select a systematic sample?	10
(b)	Show that the variance of the mean of the systematic sample is $$V_{(\bar{y}sy)} = \left(\frac{N-1}{N}\right)S^2 - \frac{k(n-1)}{N}S^2_{wsy}$$ where $S^2_{wsy} = \frac{1}{k(n-1)}\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2$	10
Q#5(a)	Derive the following expression to compare the separate and combined ratio estimates: $$V(\hat{Y}_{RC}) - V(\hat{Y}_{RS}) = \sum_{h}^{L}\frac{N_h^2(1-f_h)}{n_h}[(R-R_h)^2S^2_{x_h} + 2(R_h-R)(\rho_h S_{x_h}S_{y_h} - R_h S^2_{x_h})]$$	08

PTO

(b)	Show that the co-efficient of variation of $\hat{R}$, $\hat{\bar{Y}}_R$, $\hat{\bar{Y}}_R$ and $\hat{N}$ all equal.	12
Q#6 (a)	Show that the value of b_0 that minimizes $V(\ldots)$ is $B =$	08
(b)	Suppose that the finite population values y_i ($i = 1,2, \ldots N$) are randomly drawn from an infinite super-population in which $y_i = \alpha + \beta x_i + \varepsilon_i$ where ε_i are independent with zero mean and variance σ_ε^2 for fixed x, prove that, for any size of sample, the linear regression estimator $\bar{y}_{lr}$ is model unbiased with variance $V(\bar{y}_{lr}) = \sigma_\varepsilon^2\left[\left(\frac{1}{n} - \frac{1}{N}\right) + \frac{(\bar{X} - \ldots)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]$.	12
Q#7 (a)	Define cluster sampling. What are the reasons of using cluster sampling?	08
(b)	Calculate an expression for the optimum value of M, the size of cluster using the cost function $C = c_1 nM + c_2\sqrt{n}$ Where C is the total cost of the survey.	12
Q#8 (a)	Explain briefly how you would select a PPS sample using Hansen-Hurwitz scheme.	
(b)	An initial random sample of size n' is selected without replacement and information of x is collected. Second sample of size n is taken without replacement from the initial sample and y is measured. k is a good guess of the ratio of y to x in the population. Show that $\hat{\mu} = \bar{y} - k\bar{x} + k\bar{x}'$ is unbiased estimate of $\bar{Y}$ and $V(\hat{\mu}) = \left(\frac{1}{n} - \frac{1}{N}\right)S_y^2 - \left(\frac{1}{n} - \frac{1}{n'}\right)kS_x(2\rho S_y - kS_x)$.	
Q#9	Write a short note on the following: i. Design Effect ii. Method of controlled selection iii. Double sampling iv. Non response error	5 each