UNIVERSITY OF THE PUNJAB

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

Roll No. ....

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

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

Q.1 a) If a website receives 90 hits an hour, what is the probability *i)* they will go at least 4 minutes between hits? *ii)* we are waiting more than 2 minutes for a hit we have already waited 1 minute? *iii)* there is exactly three hits in 1 minute?

b) If the probability is 0.40 that a child exposed to a certain contagious disease will catch it, what is the probability that the tenth child exposed to the disease will be the third to catch it?

c) A graduate statistics course has 7 males and 3 female students. The professor wants to select two students at random to help her conduct a research project. What is the probability that the two students chosen are females?

(12+4+4)

Q.2 a) The amount of time a bank teller spends with each customer is normally distributed with mean 3.10 minutes and standard deviation 0.40 minutes. If a random sample of 16 customers is selected, i) what is the probability that the mean time spent per customer will be at least 3 minutes ii) there is 85% chance that the sample mean will be below how many minutes?

b) A marketing research analyst collects data for a random sample of 100 customers who purchased a particular coupon special. They spent an average of 1228.5 rupees in the store with a standard deviation of 330 rupees. Before seeing these sample results, the marketing manager had claimed that the average purchase by those responding to the coupon offer would be at least 1250 rupees. Test the claim using 5% level of significance.

c) An airline wants to estimate the proportion of non-business passengers on a new route. A random sample of 347 passengers on this route is selected, and 201 are found to be business travelers. Give a 90% confidence interval for the proportion of non-business travelers on the new route of air line.

(9+6+5)

Q.3 a) Consider the battery life of two brands of laptop computer in hours. Test at 5% level of significance that the battery life of brand-B has a larger variance than that of brand-A.

Brand-A	3.2	3.4	2.8	3.0	3.0	3.0	2.8	2.9	3.0	3.0
Brand-B	3.0	3.5	2.9	3.1	2.3	2.0	3.0	2.9	3.0	4.1

b) Find 90% confidence interval for the ratio of population variances of brand-A and brand-B.

(15+5)

Q.4 a) Fit a least square regression line for the observations of the yield of a chemical reaction taken at various temperatures.

b) Test $\beta = 0$ using t-test at 1% level of significance.

c) Test for lack of fit.

X (in degree C)	Y (%)		
150	77.4	76.7	78.2
200	84.1	84.5	83.7
250	88.9	89.2	89.7
300	94.8	94.7	95.9

(5+5+10)

P.T.O.

Q.5 a) State the properties of an F-distribution.

b) Consider the following set of observations each of which is randomly drawn from normal populations (G1, G2, G3, G4). Use Bartlett's test to test for the equality of variances.

G1	49	44	46	105	37	82
G2	97	3	69	92	106	
G3	175	95	143	200		
G4	111	178	150	83		

(5+15)

Q.6 a) A machine is adjusted to dispense acrylic paint thinner into a container. Would you say that the amount of paint thinner being dispensed by this machine varies randomly if the contents of the next 15 containers are measured and found to be 3.6, 3.9, 4.1, 3.6, 3.8, 3.7, 3.4, 4.0, 3.8, 4.1, 3.9, 4.0, 3.8, 4.2 and 4.1 litres?

b) The nicotine content (in milligrams) of two brands of cigarettes was found to be as follows. Test the hypothesis that the median nicotine contents of the two brands are equal.

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

(10+10)

Q.7 a) Prove the Brandt and Snedecor formula for chi-square.

b) A doctor believes that the proportions of births in this country on each day of the week are equal. A simple random sample of 700 births from a recent year is selected, and the results are presented in the table below. At a significance level of 0.01, is there enough evidence to support the claim of doctor?

Day	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Frequency	65	103	114	116	115	112	75

(10+10)

Q.8 a) Find $\Delta^3 y$, where $y = ax^3 + bx^2 + cx + d$ and the interval of differencing is h.

b) Consider the data below and evaluate $\sqrt{155}$ using Lagrange's interpolation formula.

x	150	152	154	156
$y = \sqrt{x}$	12.247	12.329	12.410	12.490

(10+10)

Q.9 a) What is meant by sequential analysis? Explain the general testing procedure of sequential test.

b) The results of a random sample of children with pain from musculoskeletal injuries treated with acetaminophen, ibuprofen, or codeine are shown in the table. At $\alpha = 0.1$, is there enough evidence to conclude that the treatment and results are independent?

	Acetaminophen	Ibuprofen	Codeine
Significant Improvement	58	81	61
Slight Improvement	42	19	39

(8+12)

UNIVERSITY OF THE PUNJAB

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

Roll No. ....

Subject: Statistics **Paper: II (Probability and Probability Distributions)** **Time: 3 Hrs.** **Marks: 100**

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

Q.1.a) Define the following terms: **(10)**
- i) Event space and its properties
- ii) Collectively exhaustive events
- iii) Conditional Probability
- iv) Characteristic function
- v) Random experiment

b) If 8 married couples are seated at random at a round table, compute the probability that no wife sits next to her husband. **(7)**

c) A problem is given to three persons P, Q, R whose respective chances of solving it are 2/7, 4/7, 4/9, respectively. What is the probability that the problem is solved? **(4)**

d) How many six digits numbers are there? How many of them contain the digit 4? **(4)**

Q.2.a) Bowl B_1 contains two red and four white chips, bowl B_2 contains one red and two white chips, and bowl B_3 contains five red and four white chips. The probabilities for selecting the bowls B_1, B_2 and B_3 are $\frac{1}{3}, \frac{1}{6}$ and $\frac{1}{2}$ respectively. A chip drawn at random from one of the bowls turns out to be red. Find the probabilities that it came from bowl B_1, bowl B_2 and bowl B_3. **(7)**

b) From a group of 8 women and 6 men a committee consisting of 3 men and 3 women is to be formed. How many different committees are possible if **(6)**
- i) 4 of the men refuse to serve together
- ii) 1 man and 1 woman refuse to serve together?.

c) Show that for the rectangular distribution **(6)**

$$f(x) = \frac{1}{2a} \qquad -a < x < a$$

The m.g.f. about origin is $\frac{1}{at} Sinhat$. Also show that the moments of even order given by

$$\mu_{2n} = \frac{a^{2n}}{(2n+1)}.$$

d) State and prove Chebyshev's inequality. **(6)**

Q.3.a) Derive mean and variance of Hypergeometric distribution. **(10)**

P.T.O.

b) Find the probability generating function and moment generating function of the Poisson distribution and show that its cumulants are equal. **(10)**

c) For binomial distribution, prove the relation **(5)**

$$\mu_{r+1} = pq\left(nru_{r-1} + \frac{d\mu_r}{dp}\right), \text{ where } \mu_r = rth \text{ moment about mean.}$$

Q.4 a) Derive the characteristic function of the Cauchy Distribution. **(10)**

b) Show that when n is very large and neither p nor q is very small, normal distribution is a limiting form of binomial distribution. **(10)**

c) Derive mean and variance of Weibull distribution. **(5)**

Q.5 a) Derive the χ^2-distribution. **(15)**

b) Given the joint density function **(10)**

$$f(x,y) = \begin{cases} \dfrac{x(1+3y^2)}{4} & 0 < x < 2,\ 0 < y < 1 \\ 0 & elsewhere \end{cases}$$

i) Show that the p.d.f. above is a valid density function
ii) Find marginal density of X and Y
iii) Compute E(X), Var(X) and E(XY)
iv) Find f(x/y) and evaluate $P(1/4 < X < 1/2 / Y = 1/3)$

Q.6: a) If X and Y are independent Gamma variates find the distribution of X+Y and $\dfrac{X}{X+Y}$ Where X~ gamma(α,1) and Y~ gamma(β,1). **(10)**

b) If X has a bivariate normal distribution then show that conditional distribution of Y given X = x is normal with mean $\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_x)$ and variance $\sigma_Y^2(1-\rho^2)$. **(10)**

c) If X has standard normal distribution. Find the distribution of $Y = X^2$. **(5)**

Q.7 a) A sample of odd size $n = 2r+1$ is taken from the rectangular distribution $f(x) = 1 \quad 0 \le x \le 1$ **(10)**
The median is $(r+1)th$ member of the sample arranged in ascending order. Find the variance of the median.

b) If X and Y are random variables with p.d.f **(10)**
$f(x,y) = 4xye^{-(x^2+y^2)} 0 \le x; y \le \infty$
Find the p.f of Z where $Z = \sqrt{X^2 + Y^2}$.

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

UNIVERSITY OF THE PUNJAB

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

Roll No. ....

Subject: Statistics **Paper: III (Design and Analysis of Experiments)** **Time: 3 Hrs.** **Marks: 100**

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

Q.1 a) Discuss why using two-sample t-tests is not an appropriate alternative to analysis of variance. (10)

b) Show mathematically how the total sum of squares is partitioned into the error sum of squares and treatment sum of squares. Find the number of degrees of freedom associated with each of these. (15)

Q.2 a) For the following model (12)

$$Y_{ij} = \mu + \alpha_j + \beta_i + \varepsilon_{ij}$$

$$For\ i = 1,2, \ldots r \qquad j = 1,2, \ldots .k$$

Find the expected mean squares when α and β are considered to be random.

b) Derive formula for estimating N missing values in Latin Square Design when values are missing in different columns, rows and treatments. Also deduce it to the case of two missing observations. What changes will occur in ANOVA table after estimating missing observations. (13)

Q.3 a) Given the following abbreviated analysis of variance for a Randomized Complete Block design: (15)

Source of Variation	d.f.	Sum of Squares	Mean Square
Blocks	-	0.4074	
Treatments	3	-	0.39953
Error	27	0.6249	

i. Complete the analysis after filling the missing values.
ii. Compute the standard error for a treatment mean and for the difference between 2 treatment means
iii. The treatment means are 1.464, 1.195, 1.325 and 1.662. What mean or means do you suspect might represent different population?
iv. Estimate the efficiency of this design relative to Completely Randomized design.

b) Construct a 5x5 Graeco-Latin Square experimental design, indicating the steps in the construction, and give the analysis of variance appropriate to this design. (10)

Q.4 a) What are the assumptions behind a covariance analysis? In the process of analyzing data by a covariance analysis, what tests of significance are made? (10)

P.T.O.

b) For an experiment, the results are summarized by the following sum of squares and products: (15)

Source of Variation	d.f	$\sum X^2$	$\sum Y^2$	$\sum XY$
Blocks	5	7472.6	6.31	-111.65
Treatments	6	116020.3	112.86	3598.05
Error	30	28665.1	23.23	682.20
Total	41	152158.0	142.40	4168.60

i. Based on the error sums of squares and products, is the regression of Y on X significant at $\alpha=0.05$?

ii. If yes or not compute the analysis and draw conclusion about the significance of treatment means.

Q.5 a) Write the difference between Yates technique and sign table method for computing contrasts by giving examples. (10)

b) An experiment was laid to see the effect of three fertilizers n, p and k at two levels each (i.e., applied and not applied). The eight combinations were randomized in each replication. Set up an analysis of variance to test the significance of Main Effects and Interactions from the data given below: (15)

Replication	Treatments							
	(1)	N	p	Np	k	nk	pk	npk
1	24	25	24	24	29	30	22	32
2	30	31	31	27	39	34	25	23
3	23	33	27	23	31	29	24	37
4	28	26	24	27	36	36	29	34

Q.6 a) What do you know by the term confounding? Also explain in detail what is partial confounding? (10)

b) Complete ANOVA table for the following factorial experiment. ABC is completely confounded (15)

Replicate I		Replicate II	
BI	BII	BI	BII
(1)= 10	a=6	ab= 8	a= 6
ab= 7	b= 4	(1)= 12	abc= 8
ac= 6	c= 9	ac= 7	b= 6
bc= 8	abc=5	bc= 6	c= 7

Q.7 Write the note on following topics (25)

a) Randomization
b) Degrees of freedom
c) Complete Confounding
d) Steps in construction of experimental design
e) Fixed effect and Random effect model

UNIVERSITY OF THE PUNJAB

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

Roll No. ....

Subject: Statistics **Paper: IV (Sampling Techniques)** **Time: 3 Hrs.** **Marks: 60**

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

Q#1 (a)	Explain the following concepts: i. Simple random sampling for proportions ii. Accuracy and precision with reference to sampling theory	08
(b)	Discuss the various steps involved during the planning and execution of a survey.	12
Q#2(a)	For simple random sampling, show that the sample mean $\bar{y}$ is an unbiased estimator of the population mean $\bar{Y}$ and variance of sample mean is given by $$V(\bar{y}) = \left(\frac{N-1}{Nn}\right)S^2$$ Note: Use Cornfield approach to prove this.	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)	In stratified random sampling with a linear cost function of the form $C = C_0 + \sum C_h n_h$, show that the variance of the estimated mean $\bar{y}_{st}$ is a minimum for a specified cost C, and the cost is minimum for the specified variance $V(\bar{y}_{st})$, when n_h is proportional to $\frac{W_h S_h}{\sqrt{C_h}}$.	12
(b)	Define "Preferred Samples" and "Non-preferred Samples". How the selection of sample is controlled?	08
Q#4 (a)	Describe a practical example in which cluster sampling can be used.	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)	Discuss three methods of selecting a sample that lead to an unbiased ratio type estimate.	12
(b)	Obtain the variance of $\hat{Y}_R$ which is minimum, under the model $y_i = \beta x_i + \varepsilon_i$	08
Q#6 (a)	Derive the expressions for separate and combined regression estimates of mean and variance in stratified sampling, which has the smaller variance.	10

P.T.O.

(b)	Suppose that the finite population values y_i $(i = 1,2, \ldots\ldots, N)$ 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 $V(\bar{y}_{lr}) = \sigma_\varepsilon^2 \left[\left(\frac{1}{n} - \frac{1}{N}\right) + \frac{(\bar{X}-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]$.	10
Q#7 (a)	Discuss the application of cluster sampling by the help of a practical situation.	08
(b)	If a simple random sample of n clusters, each containing M elements, is drawn from N clusters in the population, show that the sample mean per element $\bar{\bar{y}}$ is an unbiased estimate of $\bar{\bar{Y}}$ with variance: $$V(\bar{\bar{y}}) = \frac{1-f}{n}\frac{(NM-1)}{M^2(N-1)} S^2[1 + (M-1)\rho].$$	12
Q#8 (a)	What is Double sampling? Explain by the help of an example.	10
(b)	If a sample of "n" units are drawn with probability proportional to z_i and with replacement, show that an unbiased estimate of $V(\hat{Y}_{PPZ})$ is $$v(\hat{Y}_{PPZ}) = \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\frac{y_i}{z_i} - \hat{Y}_{PPZ}\right)^2$$	10
Q#9	Write a short note on the following: i. Circular systematic sampling ii. Inverse sampling iii. Two-stage sampling iv. Non-response error	5 each