UNIVERSITY OF THE PUNJAB

M.A./M.Sc. Part – I Supply – 2020 & Annual – 2021

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) The probability that a student pilot passes the written test for a private pilot's license is 0.7. Find the probability that the student will pass the test
   i. On the third try
   ii. Before the fourth try **(10)**

b) The probability that a person, living in a certain city, owns a dog is estimated to be 0.3. Find the probability that the tenth person randomly interviewed in that cite is the fifth one to own a dog. **(10)**

Q.2.

a) Discuss the relationship between geometric distribution and negative binomial distribution. **(6)**

b) Suppose that, on average, 1 person in 1000 makes a numerical error in preparing his or her income tax return. If 10,000 forms are selected at random and examined, find the probability that 6, 7, or 8 of the forms contain an error. **(6)**

c) An annexation suit is being considered against a county subdivision of 1200 residences by a neighboring city. If the occupants of the half of the residences object to being annexed, what is the probability that in a random sample of 10 at least 3 favor the annexation suit? **(8)**

Q.3.

a) The average grade for an exam is 74, and the standard deviation is 7. If 12% of the class is given A's, and the grades are curved to follow a normal distribution, what is the lowest possible A and the highest possible B? **(8)**

b) Define P-value **(5)**

c) An electrical firm manufactures light bulbs that have a lifetime that is approximately normally distributed with a mean of 800 hours and a standard deviation of 40 hours. Test the hypothesis that μ=800 hours against the two sided alternative if a random sample of 30 bulbs has an average life of 788 hours. Use *P*-value in your conclusion. **(7)**

Q.4.

a) Complete the following ANOVA table:

Source of variation	Sum of squares	Degrees of freedom	Mean square	Computed f	P-values	Decision
Regression	509.2507	1			0.0001	
Error						
Lack of fit	1.2060	2			0.2241	
Pure error						
Total	513.1167	11				

(10)

b) Test the hypothesis that $\beta = 0$ at 1% level of significance by using the given data:

X	70	92	80	74	65	83
Y	74	84	63	87	78	90

(10)

P.T.O.

Q.5.

a) The grades in Statistics course for a particular semester were as follows:

Grade	A	B	C	D	E
F	14	18	32	20	16

Test the hypothesis at 5% level of significance that the distribution of grades is uniform. **(10)**

b) A builder claims that the heat pumps are installed in 70% of all homes being constructed today in the city. Would you agree with his claim if a random survey of new homes in this city shows that 8 out of 15 had heat pumps installed? Use a 0.10 level of significance. **(10)**

Q.6.

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

Brand A	2.1	4	6.3	5.4	4.8	3.7	6.1	3.3		
Brand B	4.1	0.6	3.1	2.5	4	6.2	1.6	2.2	1.9	5.4

Test the hypothesis, at 5% level of significance, that the median nicotine contents of the two brands are equal against the alternative that they are unequal. **(10)**

b) Explain the general procedure of the Run test for one sample. **(5)**

c) The weights of 5 people before they stopped smoking and 5 weeks after they stopped smoking, in kilograms, are as follows:

	1	2	3	4	5
Before	66	80	69	52	75
After	71	82	68	56	73

Use the signed rank test for paired observations to test the hypothesis, at 5% level of significance, that giving up smoking has no effect on a person's weight against the alternative that one's weight increases if he or she quite smoking. **(5)**

Q.7.

a) ~~Explain the difference between parametric and non-parametric tests.~~ **(8)**

b) The following data represents the number of hours that a rechargeable hedge trimmer operated before a recharge is required:

1.5	2.2	0.9	1.3	2.0	1.6	1.8	1.5	2.0	1.2	1.7

Use an appropriate test to test the hypothesis at 5% level of significance that this trimmer operates with a median of 1.8 hours before requiring a recharge. **(6)**

c) 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	3.8	4.2	4.1	4.0	

Use a 0.1 level of significance. **(6)**

Q.8.

a) Compute the third difference of f(51) by the formula $\Delta^n f(x) = (E-1)^n f(x)$ from the following entries, verify your results by means of a difference table.

X	51	52	53	54
F(x)	132650	140608	148877	157464

(10)

b) Show that the first difference of the linear function and the second difference of the quadratic is constant. **(10)**

Q.9.

a) Apply Lagrange's formula inversely to find the value of x for which f(x)=50, from the following data:

x	14	17	31	35
$f(x)$	68.7	64.0	44.0	39.1

(12)

b) Explain the general testing procedure of sequential test. **(8)**

UNIVERSITY OF THE PUNJAB

M.A./M.Sc. Part – I Supply – 2020 & Annual – 2021

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) Write short notes on **(10)**

i) Event space
ii) Exhaustive events
iii) Cumulant generating function
iv) Memory less property
v) Distribution function

b) A blood test is 99% effective in detecting a certain disease when the disease is present. However, the test also yields a false-positive result for 2 % of the healthy patients tested (that is, if a healthy person is tested, then with probability 0.02 the test will say that this person has the disease). Suppose 0.5% of the population has the disease. Find the probability that a randomly tested individual actually has the disease given that his or her test result is positive. **(06)**

c) A number is selected at random from the set of natural numbers {1,2,3,4....1000}. What is the probability that it is not divisible by 4, 7 or 9? **(05)**

d) Albert goes to the grocery store to buy fruits. There are 7 different varieties of fruits and Albert is determined to buy no more than one of any variety. How many different orders can he place **(04)**

Q.2.a) State and prove the Bayes Theorem. **(08)**

b) A town has 2 fire engines operating independently. The probability that a specific engine is available when needed is 0.96. **(06)**

i) What is the probability that neither is available when needed?
ii) What is the probability that a fire engine is available when needed?

c) A die is loaded in such a way that each odd number is twice as likely to occur as each even number. What is the probability that **(06)**

(i) A number greater than 3 occurs on a single roll of the die?
(ii) The number of points rolled is a perfect square?
(iii) It is perfect square given that it is greater than 3?

d) Twenty percent (20%) of a city's residents are in favor of complete lock down to slow down the spread of COVID-19, 63% are against it, and 17% are indifferent (having no opinion). What is the probability that a randomly selected person from this city will either be against it or be indifferent? What is the type of the events? **(05)**

Q.3.a) Obtain the first four factorial cumulants and first four mean moments of the binomial distribution and hence show that (10)

$$\gamma_1 = \frac{1-2p}{\sqrt{npq}} \text{ and } \gamma_2 = \frac{1-6pq}{npq}$$

b) Find β_1 and β_2 for the Geometric distribution. (8)

c) If the p.d.f. of X is given by (7)

$$f(x) = 630x^4(1-x)^4 \quad 0 < x < 1$$
$$0 \quad \text{otherwise}$$

Find the probability that it will take a value within two standard deviations of the mean and compare this probability with the lower bound provided by Chebyshev's inequality.

Q.4.a) Show that the mean and standard deviation of exponential distribution are equal Also find the mean deviation of exponential distribution. (10)

b) Derive Mean and Variance of Weibull distribution. (8)

c) Show that if X follows a t-distribution with ν d.f, then X^2follows the F-distribution with $F(1,\nu)$. (7)

Q.5.a) Obtain the r^{th} moment about origin of F-distribution. Also find its mean and variance. (8)

b) Let X_1 and X_2 be independent variable with " 0 " mean and unit variance, then obtain the distribution of $Z = X_1 / X_2$. Also name it. (10)

c) Derive the Chi-square distribution. (7)

Q.6.a) If X and Y are independent Gamma variates. Find the distribution of $X+Y$ and $\frac{X}{X+Y}$. where $X \sim \gamma(\alpha,1)$ and $Y \sim \gamma(\beta,1)$. (10)

b) If (x, y) has a bivariate normal distribution then show that the marginal distribution of X is univariate normal distribution i.e. $X \sim N(\mu_1, \sigma_1^2)$ (8)

c) Let X and Y be independent random variables with joint p.d.f. (7)

$$f(x, y) = x + y \quad 0 < x < 1,\ 0 < y < 1$$
$$= 0 \quad \text{elsewhere}$$

Find E(XY) and E(X + Y).

Q.7.a) Derive the distribution of rth order statistic. (10)

b) Show that in a sample of n observation from $f(x) = \theta^{-x};\ 0 \le x \le \infty$ the variance of smallest observation is $\frac{1}{n^2}$. (10)

c) Find moment generating function for the normal distribution. (05)

UNIVERSITY OF THE PUNJAB

M.A./M.Sc. Part – I Supply – 2020 & Annual – 2021

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) When we use multiple comparison tests? Explain the use of Duncan's Multiple Range (DMR) test.

b) Four varieties of wheat are compared in five randomized complete blocks. Following computations are made: Total SS=182.17, Error SS=26.26 and Varieties SS=134.45.

(i) Test the hypothesis that all four varieties have same average yield.

(ii) Compare the efficiency of this design with the completely randomized design if blocks were ignored. (15+10)

Q.2 a) Derive formula for estimating N missing values in a Latin Square (LS) Design when values are missing in different columns, rows and treatments. Deduce this formula when two values are missing.

b) For a Randomized Complete Block experiment

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \qquad for\ i = 1,2,...,a.and\ j = 1,2,...,b.$$

Let $\hat{\mu}, \hat{\tau}\ and\ \hat{\beta}$ are least square estimates of μ, α_i and β_j respectively. Develop expected Mean Squares indicating the assumptions.

c) What is a Graeco Latin Square Design? Construct a Graeco to study the effects of five treatments. Outline the ANOVA table. (9+4+12)

Q.3 a) A big car company uses four assembly methods (A, B, C, D) for its cars. An industrial engineer wants to investigate the effect of these four methods on the assembly time of a car. Four operatures are selected for the study. The engineer knows that each assembly method produces such fatigue that the time required for the last assembly may be different from the time required for the first, regardless of the method. That is a trend develops in the required assembly time. Analyze the data from this experiment at α=0.01 and draw appropriate conclusion.

Order of Assembly	Operator			
	1	2	3	4
1	$C = 10$	$D = 14$	$A = 7$	$B = 8$
2	$B = 7$	$C = 18$	$D = 11$	$A = 8$
3	$A = 5$	$B = 10$	$C = 11$	$D = 9$
4	$D = 10$	$A = 10$	$B = 12$	$C = 14$

b) In an RCB design with *p* treatments with *a* observations yield is assumed to be represented by the model

$$Y_{ij} = \mu + \alpha_i + \gamma_j + \beta(X_{ij} - \overline{X}) + \varepsilon_{ij}$$

Develop procedure for testing the null hypothesis that the adjusted treatment means are equal. (12+13)

Q. 4 a) What are the useful properties of a factorial design?

b) Given below are the totals of response for different treatment combinations with 3 replications of a three factor factorial experiment.

(I)	a	b	ab	c	ac	bc	abc
100.7	117.2	103.8	144.3	100.9	106.2	98.5	142.8

Complete the ANOVA table and draw conclusions.

c) Compute standard error of treatment mean. (9+10+6)

Q.5 a) What is blocking? What do you mean by confounding? Differentiate between blocking and fractionalization of factorial eperiments.

b) Complete ANOVA table for the following factorial experiment. ABC is confounded in Replicate I and AB in Replicate II.

Rep.I.		Rep. II	
1	2	1	2
b = -1	(1)= -3	(l) = -1	b = 0
a= 0	ab= 2	abc =5	ac =1
c= -1	bc= 1	c = 0	bc = 1
abc= 6	ac= 2	ab=3	a = 1

c) Construct a 2^{6-2} factorial design with highest possible resolution. Outline the pairs of aliases and first two columns of ANOVA table. Also tell the resolution of produced design. (8+8+9)

Q.6 a) Compare and contrast 2^k and 3^k experiments. How these are different in the estimation of model?

b) In a split plot design the whole plot treatment (A) was applied to 3 blocks and treatment (B) was then applied to split plots. The total SS, Block SS, whole Plot Error SS are 822.97, 77.55 and 36.28 respectively. The sum for 3 blocks are given in the following table:

	B_1	B_2	B_3	B_4
A_1	89	104	118	117
A_2	100	117	119	126
A_3	92	92	104	121

Complete the ANOVA and draw conclusions.

c) For the above data in b) compute the standard error for the
i) difference between two 'A' treatment means
ii) difference between two 'B' treatment means (6+12+7)

Q.7 a) Write a short note on Incomplete Latin Square Design.

b) An engineer is studying the milage performance characteristics of five types of gasoline additives. In the road tests he wishes to use cars as blocks. Due to some constraints he used an incomplete block design. Analyse the data and draw conclusions.

	Cars				
Additives	1	2	3	4	5
1	-	17	14	13	12
2	14	14	-	13	10
3	12	-	13	12	9
4	13	11	11	12	-
5	11	12	10	-	8

c) Consider the following partially balanced incomplete block design

Blocks	Treatment	Combinations	
1	1	2	3
2	3	4	5
3	2	5	6
4	1	2	4
5	3	4	6
6	1	5	6

Verify the following relationships among the parameters

$p_{11}^1 + p_{12}^2 = n_1$

$n_1 p_{12}^1 = n_2 p_{11}^2$

$p_{21}^1 + p_{22}^1 = n_2$

$n_1 p_{11}^1 = n_2 p_{12}^2$

$p_{11}^1 + p_{12}^1 = n_1 - 1$

$n_1\lambda_1 + n_2\lambda_2 = r(k-1)$

$p_{21}^1 + p_{22}^2 = n_2 - 1$

$n_1 + n_2 = a - 1$

(5+10+10)

UNIVERSITY OF THE PUNJAB

M.A./M.Sc. Part – I Supply – 2020 & Annual – 2021

Roll No. .....

Subject: Statistics Paper: IV (Sampling Techniques) Time: 3 Hrs. Marks: 100

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	12
(b)	If the loss function due to an error in $\bar{y}$ is $d(\bar{y}-\bar{Y})^2$ and the cost function is $C = C_0 + C_1 n$, then show that the most economical value of 'n' in simple random sampling, ignoring finite population correction is $\sqrt{\frac{dS^2}{C_1}}$.	08
Q#2(a)	What is Design Effect (Deff)? Explain briefly.	08
(b)	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 minimum for specified cost C and cost is minimum for specified variance $V(\bar{y}_{st})$ when $n_h \propto \frac{W_h S_h}{\sqrt{C_h}}$.	12
Q#3 (a)	Describe the idea of inverse sampling in detail.	10
(b)	Prove that sample proportion is an unbiased estimator of population proportion with variance $$V_{pst} = \frac{1}{N^2}\sum \frac{N_h^2(N_h - n_h)}{N_h - 1}\frac{P_h Q_h}{n_h}$$	10
Q#4(a)	Contrast between systematic sample and other selection procedures. If i) the population is linear ii) the population is in random order	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)	Compare regression estimation with ratio estimation and mean per unit	10
(b)	Obtain the condition upon ρ for which $V(\hat{\bar{Y}}_R)$ is smaller than $V(\hat{\bar{Y}})$.	10

P.T.O.

Q#6 (a)	If b is the least square estimate of B and $\bar{y}_{lr} = \bar{y} + b(\bar{X} - \bar{x})$, then prove that for simple random sampling of size n (assuming large), the resulting minimum variance is $$V(\bar{y}_{lr}) = \frac{1-f}{n} S_y^2(1-\rho^2)$$ Where ρ is the population correlation coefficient between y and x.	10
(b)	Show that the value of b_0 that minimizes $V(\bar{y}_{lr})$ is $B = \frac{S_{yx}}{S_x^2}$.	10
Q#7(a)	Define cluster sampling. What are the reasons of using cluster sampling?	08
(b)	A simple random sample of n clusters, each containing M elements, is drawn from the N clusters in the population. Then the sample mean per element $\bar{y}$ is an unbiased estimate of $\bar{\bar{Y}}$ with variance $V(\bar{y}) = \frac{1-f}{nM} S^2[1 + (M-1)\rho]$	12
Q#8 (a)	What is non-response error? Also describe the sources of non-response.	08
(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 $V(\hat{\mu}) = \left(\frac{1}{n} - \frac{1}{N}\right) S_y^2 - \left(\frac{1}{n} - \frac{1}{n'}\right) kS_x(2\rho S_y - kS_x)$.	12
Q#9	Write a short note on the following: i. Optimum Allocations in stratified random sampling ii. Advantages of systematic sampling iii. Two stage sampling iv. Double Sampling	5 each