UNIVERSITY OF THE PUNJAB

M.A./M.Sc. Part – I Annual Examination – 2022

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 number of people arriving at a fast-food drive-thru in any given 2-minute interval obeys a Poisson process with a mean of 1. Suppose that the waiters can only process 3 orders in any given 4-minute interval (also, assume the waiters can process an order instantaneously but are limited in how many they can process in given time intervals, as indicated). What is the expected number of people that leave the drive-thru with their orders filled in any given 4-minute interval? **(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) A large agricultural producer in Spain produces melons with diameters that are normally distributed with a mean of $\mu = 15$ centimeters (cm) and a standard deviation of $\sigma = 2$ cm. Find the solutions to these questions.

i. What is the probability a melon will have a diameter of at least 12 cm?

ii. What is the probability that a randomly selected melon will have a diameter of no less than 12 cm but no more than 16 cm?

iii. The producer has an arrangement with several gourmet shops by which it will receive a slightly higher price for melons with a diameter that falls in the top 10%. What is the minimum diameter a melon must have in order to qualify for the higher price? **(10)**

b) On the average, a certain computer part lasts ten years. The length of time the computer part lasts is exponentially distributed. What is the probability that a computer part lasts more than 7 years? **(10)**

Q.3. a) In a random sample of 1000 homes in a certain city, it is found that 228 are heated by oil. Find the 99% confidence interval for the proportion of homes in this city that are heated by oil. **(10)**

b) A manufacturer of car batteries claims that the life of his batteries is approximately normally distributed with a standard deviation equal to 0.9 years. If a random sample of 10 of these batteries has a standard deviation of 1.2 years, do you think that $\sigma > 0.9$ years? Use a 0.05 level of significance. **(10)**

Q.4. a) 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 u=800 against a two sided alternative if a random sample of 30 bulbs has an average life of 788 hours. Use a P-value in your conclusion. **(10)**

b) A dry cleaner establishment claims that a new spot remover will remove more than 70% of the spots to which it is applied. To check this claim, the spot remover will be used on 12 spots chosen at random. If fewer than 11 of the spots are removed, we shall not reject the null hypothesis that $p=0.7$, otherwise we conclude that $p>0.7$.

i. Evaluate α, assuming that $p=0.7$

ii. Evaluate β, for the alternative $p=0.9$

iii. Compute the power of the test for $p=0.9$ **(10)**

Q.5. a) Compute an interpret the correlation coefficient, coefficient of determination for the following grades of 6 students selected at random:

Mathematics grades	70	92	80	74	65	83
English grades	74	84	63	87	78	90

(10)

b) Observations of the yield of a chemical reaction taken at various temperatures were recorded and provided in the table given below. Estimate the linear model $U_{y/x} = \alpha + \beta X$ and test for lack of fit. **(10)**

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

Q.6.

a) The estimated regression line computed from 32 pairs of Y and X is $y = 3.83 + 0.904X$ with, $SSE = 323.3$ *and* TSS=3713.9.

i. Draw an ANOVA table.

ii. Test the hypothesis that X and Y are independent. **(5)**

b) Discuss the assumptions and properties of the least square regression line. **(10)**

c) Differentiate between partial correlation coefficient and multiple correlation coefficient. **(5)**

Q.7.

a) In an experiment to determine which of three different missiles systems is preferable, the propellant burning rate is measured. The data after coding are given in the table. Using 5% level of significance test the hypothesis that the propellant burning rates are same for the three missile systems. **(10)**

Missile systems		
24.0, 22.8, 16.7 19.8, 18.9	23.2, 19.8, 18.1 17.6, 20.2, 17.8	18.4, 19.1 17.3 17.3 19.7 18.9 18.8 19.3

b) A cigarette manufacturer claims that the tar content of brand B cigarettes is lower than that of brand A. To test this claim, the following determinations of tar content, in milligrams, were recorded:

Brand A	1	12	9	13	11	14
Brand B	8	10	7			

Use the rank-sum test with $\alpha = 5\%$ to test whether the claim is valid. **(7)**

c) Explain the term "Distribution free methods". **(3)**

Q.8.

a) Define interpolation. **(3)**

b) The following values of U_x are given:

$U_{10}=544$, $U_{15}=1227$, $U_{20}=1775$

Find, correct to one decimal place, the value of x for which $U_x=1000$. **(6)**

c) Given the table of values

X	-3	-2	-1	0	1
Y	16	7	4	1	-8

Find by means of Gregory-Newton formula, an expression of Y as a function of X. **(5)**

d) Express $f(a+nh)$ in terms of $f(a)$ and successive differences of $f(a)$, where n is a positive integer. **(6)**

Q.9.

a) Explain how you would test the equality of k (k>2) variances for unequal sample sizes. **(5)**

b) Prove that $\chi^2 = \dfrac{(ad-bc)^2(a+b+c+d)}{(a+b)(c+d)(a+c)(b+d)}$ **(10)**

c) Discuss the significance of the sequential sampling. **(5)**

UNIVERSITY OF THE PUNJAB

M.A./M.Sc. Part – I Annual Examination – 2022

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
- ii) Indicator function
- iii) Law of large numbers
- iv) Factorial moments
- v) Conditional probability

b) Five people designated as A, B, C, D and E arranged in a row, assuming selection of the persons is equally likely, what is the probability that **(10)**
i) there is exactly one person between A and B?
ii) there are exactly two persons between A and B?
iii) there are exactly three persons between A and B?
iv) A and B must sit next to each other?
v) B and C must not sit next to each other?

c) If X is the number of tails obtained in four tosses of a balanced coin. Find the probability distribution of $Z=(X-1)^2$. **(5)**

Q.2.a) A student has selected 3 courses A, B and C to study in a semester. His chances of pass in these courses are 0.8, 0.6, and 0.9 respectively. Find the probability that he will pass at least one course. Also name the probability law to be used in this case. **(8)**

b) Police plan to enforce speed limits by using radar traps at 4 different locations within the city limits. The radar traps at each of locations L1, L2, L3, and L4 are operated 40%, 30%, 20%, and 30% of the time, and if a person who is speeding on his way to work has probabilities of 0.2, 0.1, 0.5, and 0.2 respectively, of passing through these locations, what is the probability that he will receive a speeding ticket? If the person received a speeding ticket on his way to work, what is the probability that he passed through the radar trap located at L2? **(10)**

c) In a certain town, 40% of the eligible voters prefer candidate A, 10%, Prefer candidate B, and the remaining 50% have no preference. You randomly sample 10 eligible voters. What is the probability that 4 will prefer candidate A, 1 will prefer candidate B, and the remaining 5 will have no preference? **(7)**

Q.3.a) Obtain the first four factorial cumulants and first four mean moments of the binomial distribution and hence show that **(12)**

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

b) Prove that when N becomes indefinitely large, the Hyper-geometric distribution tends to the Binomial distribution. **(8)**

c) Prove that the geometric distribution has the memory less property. **(5)**

Q.4.a) Show that the mean and standard deviation for a variate following the exponential distribution are equal. Also find the mean deviation of for such a variate. **(9)**

b) Find mean, variance and mode of Beta distribution of the first kind. **(9)**

c) If the probability density of X is given by **(7)**

$$f(x) = 630\, x^4 (1-x)^4 \qquad 0 < x < 1$$

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

Q.5.a) Derive the F-distribution. **(10)**

b) Show that mean, median and mode of normal distribution are identical. **(7)**

c) Find μ_r' for the Pareto distribution given by **(8)**

$f(x;k,a) = \frac{ak^a}{x^{a+1}}; a > 0, k > 0, x \geq k$ and zero for x < k

Hence find mean and variance.

Q.6.a) 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 $\sigma_Y^2(1-\rho^2)$. **(9)**

b) Three fair coins are tossed let X denote the number of heads and Y denote the number of runs of heads **(9)**

i) Find the joint distribution of X and Y

ii) Find E(Y / X = 2) and V(Y / X = 2).

c) Derive limiting form of χ^2-distribution **(7)**

Q.7.a) Let x_1, x_2, x_3, x_4, x_5 are observations of a random sample from a continuous distribution having p.d.f. **(10)**

$f(x) = e^{-x}, 0 < x < \infty$ and if these observations are arranged in ascending order of magnitude then obtain the distribution of the median.

b) Let X and Y be independent standard normal variables. Find the pdf of Z=X / Y. **(10)**

c) Let random variable X have the pdf given by **(5)**

$$f(x) = \begin{cases} 4x^2 & 0 < x < \infty \\ 0 & otherwise \end{cases}$$

Find the distribution of Y= - log X.

UNIVERSITY OF THE PUNJAB

Roll No. …………….

M.A./M.Sc. Part – I Annual Examination – 2022

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) Why we use multiple comparison tests? How these differ in application? Explain least significant defference (LSD) test.

b) Four randomized complete blocks are formed to compare four varieties of sunflower. The Total SS, Error SS and Varieties SS are 182.17, 26.26 and 134.45 respectively.

(i) Test the significance of difference between varietal means.

(ii) Compare the efficiency of this design with the design in which blocks are ignored. (15+10)

Q.2 a) Construct 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

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

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) The effect of four assembly methods (A, B, C, D) are investigated by an industrial engineer on the assembly time for a colour TV component. 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 fist, regardless of the method. That is a trend develops in the required assembly time. Analyze the data from this experiment (α=0.05) 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 main reasons of using a factorial design?

b) Given below are the totals of response for different treatment combinations in a 2^3 experiment with 2 replications.

(l)	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? Explain the difference between complete and partial confounding with examples.

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

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

c) Construct a 2^{7-2} 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) Why we have to choose irregular factorial dsigns in place of regular factorial designs in some situations?

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 Desian.

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

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

$n_1 p_{12}^1 = n_2 p_{11}^2$ $\quad$ $n_1\lambda_1 + n_2\lambda_2 = r(k-1)$ $\quad$ (5+10+10)

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

$n_1 p_{11}^1 = n_2 p_{12}^2$ $\quad$ $n_1 + n_2 = a - 1$

UNIVERSITY OF THE PUNJAB

M.A./M.Sc. Part – I Annual Examination – 2022

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)	How do we select a random sample? Explain following two methods for random selection of sample: 1. Lottery Method 2. Random Number Tables	12
(b)	If the loss function due to an error in $\bar{y}$ is $\lambda\lvert\bar{y}-\bar{Y}\rvert$ 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 $(\frac{\lambda S}{C_1\sqrt{2\pi}})^{2/3}$	08
Q#2(a)	Describe the different ways of allocation of sample size in strata.	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)	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. Describe the limitations of systematic sampling procedure.	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 $V(\hat{Y}_R)$ is smaller than $V(\hat{Y})$.	10
Q#6 (a)	For simple random sampling in which b_0 is a pre assigned constant, show that the linear regression estimate $\bar{y}_{lr} = \bar{y} + b_0(\bar{X} - \bar{x})$ is an unbiased estimate of $\bar{Y}$ with variance $$V(\bar{y}_{lr}) = \frac{1-f}{n}(S_y^2 - 2b_0 S_{yx} + b_0^2 S_x^2)$$	10
(b)	Obtain the condition upon ρ for which $V(\hat{Y}_R)$ is smaller than $V(\hat{Y})$.	10
Q#7(a)	Define two stage cluster sampling. What are the reasons of using two stage 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 $\bar{\bar{y}}$ is an unbiased estimate of $\bar{\bar{Y}}$ with variance $V(\bar{\bar{y}}) = \frac{1-f}{nM} S^2[1 + (M-1)\rho]$	12
Q#8 (a)	What is non-response error? Also describe the sources of non-response.	08
(b)	If S_b^2 is the variance between units in the population and $S_w^2 = AM^g$ $(g > 0)$ is the variance between elements that lie in the same unit, find the optimum value of M, the size of the unit.	12
Q#9	Write a short note on the following: i. Stratified random sampling ii. Disadvantages of sampling ~~iii. Ratio estimator~~ iv. Design Effect	5 each