UNIVERSITY OF THE PUNJAB

Roll No. ....

M.A./M.Sc. Part – I Annual Examination – 2020

Time: 3 Hrs. Marks: 100

Subject: Statistics Paper: I (Statistical Methods)

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

Q.1.

a) 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)**

b) A government task force suspects that, some manufacturing companies are in violation of federal pollution regulations with regard to dumping a certain type of product. Twenty firms are under suspicion but all cannot be inspected. Suppose that 3 of the firms are in violation.

i. What is the probability that inspection of five firms find no violation? **(10)**

ii. What is the probability that plan above will find two violations?

Q.2.

a) An oil company conducts a geological study that indicates that an exploratory oil well should have a 20% chance of striking oil.

i. What is the probability that the first strike comes on the third well drilled?

ii. What is the probability that the third strike comes on the seventh well drilled?

iii. What is the mean and variance of the number of wells that must be drilled if the oil company wants to set up three producing wells? **(10)**

b) 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? **(10)**

Q.3.

a) 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)**

b) The length of time X to complete a job is exponentially distributed with mean u=10.

i. Compute the probability of job completion between two consecutive jobs exceeding 20 hours.

ii. The cost of job completion is given by $C = 4 + 2X + 2X^2$. Find the expected value of C. **(10)**

Q.4.

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. **(5)**

b) How large a sample is required if we want to be 95% confident that our estimate (p=0.68) is within 0.02? **(5)**

c) 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.5.

a) 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 $p=0.9$ **(9)**

iii. Compute the power of the test for $p=0.9$

b) Compute an interpret the correlation coefficient 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

(11)

Test the hypothesis that $\rho = 0$ by using 5% level of significance.

P.T.O.

Q.6.

a) It is of interest to study the effect of population size in various cities in US on ozone concentration. The data set consists of 1999 population in millions and the amount of the ozone present per hour in ppb (parts per billion). The data are as follows:

Y,(Ozone, ppb/hour)	X(population)	Y (%)	X (°C)
126	0.6	128	0.6
135	4.9	129	2.3
124	0.2		
128	0.5		
126	1.1		
128	2.3		

Fit the linear regression model and test for lack of fit. **(10)**

b) The estimated regression line computed from 20 pairs of Y and X is

$$y = -0.7374+0.003227x \quad \text{with, } SSE=0.1545 \text{ and } \sum_{i=1}^{20}(X_i-\bar{X})^2 = 39463.2.$$

i. Compute s_e (standard error of estimate).

ii. Test the hypothesis that X and Y are independent. **(10)**

Q.7.

a) 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 α=5% to test whether the claim is valid. **(7)**

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

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. **(10)**

Q.8.

a) Express the nth differences of tabulated function in terms of successive entries. **(5)**

b) Find the value of the argument x for which $f(x)$ is 18,600 by using the given Values

X	52	53	54	55	56
$f(x)$	19231	18868	18519	18182	17855

(10)

c) Express the relationship between the difference operator and the shift operator. **(5)**

Q.9.

a) Define the term "Interpolation". **(5)**

b) If $f(x) = g(x) + h(x)$, where $g(x) = x^4$ and $h(x) = x^3$, verify that $f(5,7,11,13) = g(5,7,11,13) + h(5,7,11,13)$. **(10)**

c) Explain how would you test the equality of k (k>2) variances. **(5)**

UNIVERSITY OF THE PUNJAB

M.A./M.Sc. Part – I Annual Examination – 2020

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 **(12)**
- i) Marginal Distribution
- ii) Moment Generating Function
- iii) Chebyshev's inequality
- iv) Random experiment

b) Eight sticks are each broken into one long and one short part. From the sixteen parts, pairs are chosen at random and new sticks are formed. Determine the probability that (I) the resulting sticks have their original form (ii) every long part is joined to a short part. **(8)**

c) One bag contains 4 white balls and 5 black balls, and a second bag contain 5 white and 5 black balls. One ball is drawn from the first bag and placed unseen in the second bag. What is the probability that a ball drawn from second bag is white? **(5)**

Q.2.a) Poker dice is played by simultaneously rolling 5 dice. Find out the probability that it contains **(8)**
i) No two alike ii) One pair iii) Two pairs iv) Full house

b) Of the travelers arriving at a small airport, 60% fly on major airlines. 30% fly on privately owned planes, and the remainder fly on commercially owned planes not belonging to a major airline. Of those traveling on major airlines, 50% are traveling for business reasons, whereas 60% of those arriving on private planes and 90% of those arriving on other commercially owned planes are traveling for business reasons. Suppose that we randomly select one person arriving at this airport. What is the probability that the person **(9)**
i) is traveling on business?
ii) arrived on a privately owned plane, given that the person is traveling for business reasons?

c) The letters of the word "TRIANGLE" are scrambled and then these letters are rearranged to form a word. How many possible arrangements are there that a word will, **(8)**
- i) Begin with T?
- ii) Begin with T or E?
- iii) Begin with T and end with E?
- iv) T and E are next to each other?

Q.3.a) Show that binomial distribution is a limiting case of Hyper-geometric distribution. **(10)**

b) Find the expected value, standard deviation, γ_1 and γ_2 of the negative binomial distribution for which **(10)**

$$P(x)=\binom{x+r-1}{x}p^r q^x \quad x=0,1,2\ldots\ldots\ldots$$

c) Find cumulative distribution function of the geometric distribution. **(5)**

Q.4.a) Derive rth moment about origin of a lognormal distribution. Hence find mean and variance for the lognormal distribution. **(10)**

b) Derive γ_1 and γ_2 of the gamma distribution. **(10)**

c) Find median of the Cauchy distribution. **(5)**

Q.5.a) Derive student's t-distribution. **(10)**

b) Toss a coin 5 times. Let X represents number of heads and Y represents the number of runs of X. Construct joint probability table of X and Y. Also, calculate correlation coefficient between X and Y. **(10)**

c) Show that the square of t-variate with "n" degree of freedom is distributed as F with "1" & "n" d.f. **(5)**

Q.6.a) Let X_1 and X_2 be independent standard normal variates, then obtain the distribution of $Z = \frac{x_1}{x_2}$. **(10)**

b) State and prove Central limit theorem. **(10)**

c) Find the distribution of $Y = X^2$, given that **(5)**

$$f(x) = \frac{1}{\pi_1(1+x^2)} \qquad -\infty \leq x \leq \infty$$

Q.7.a) Derive the joint distribution of the rth order Statistic (y_r) and the Sth order Statistic (y_s). **(10)**

b) Let $f(x) = \frac{1}{\beta} e^{-x/\beta}$, $x \geq 0$ **(10)**

Show the distribution of range is

$$h(R) = \frac{(n-1)}{\beta} e^{-R/\beta} (1 - e^{-R/\beta})^{n-2}$$

c) If $X_1, X_2, \ldots\ldots, X_n$ are independent Bernoulli r.v's $P(X_i = 1) = p$ and $P(X_i = 0) = 1 - p$. Find the p.f. of $Y = X_1 + X_2 + \ldots\ldots + X_n$. **(5)**

UNIVERSITY OF THE PUNJAB

Roll No. ....

M.A./M.Sc. Part – I Annual Examination – 2020

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) Define the following basic terms of experimental design:
i) treatment ii) experimental unit iii) experimental error
iv) correction factor v) degree of freedom (15)
Give examples.

b) Six randomized complete blocks are formed to compare four varieties of mustard. The SS for Total, Blocks and Varieties are 131.3, 69.0 and 28.2 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. (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) Consider a fixed effect model for a randomized complete block experiment

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

where $\hat{\mu}, \hat{\tau}\ and\ \hat{\beta}$ are least square estimates of μ, α_i and β_j respectively. Establish relations of expected Mean Squares of treatment and error indicating the assumptions.

c) What is a Graeco Latin Square Design (GLSD)? Construct a GLSD to study the effects of five treatments. Outline the ANOVA table. What are the advantages of using a GLSD?

(9+4+12)

Q.3 a) Seven treatments arranged in six randomized complete blocks gave the following sum of squares and products

S.O.V.	YY	XY	XX
Blocks	1200	600	200
Treatments	800	300	100
Error	1400	700	600

(i) Is the regression of Y on X significance at 5% level of significance?
(ii) Construct ANOVA table and draw inferences.

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

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

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

Q. 4 a) What is a factorial design and how it is different from the basic one-factor-at-a-time design?

b) Given below are the treatment totals for different treatment combinations in an RCB design with 3 replications.

(I)	N	P	NP	K	NK	PK	NPK
100.7	117.2	103.8	144.3	100.9	106.2	98.5	142.8

The block totals are 303.1, 308.6 and 302.7. The total sum of squares are 1031.48. Complete the ANOVA table and draw conclusions.

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

Q.5 a) 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. ABC is completely confounded with blocks.

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

c) Produce a quarter replicate of 2^6 factorial experiment choosing Two four factor effects as defining contracts. Outline the pairs of aliases and first two columns of ANOVA table. What is the resolution of this design.(8+8+9)

Q.6 a) Give the limitations of regular factorial design.

b) Complete the following ANOVA and draw conclusions.

S.O.V	d.f.	S.O.S	M.S
Replicate	3	28.44	-
A	3	-	1.64
Linear	-	4.75	-
Non-linear	-	-	-
Error(I)	-	-	-
Subtotal	-	40.85	-
B	-	29.30	-
Linear	-	-	-
Nonlinear	-	4.55	-
AB	9	-	-
Error(II)	-	86.65	-
Total	-	206.69	-

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

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

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

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

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

UNIVERSITY OF THE PUNJAB

Roll No. ....

M.A./M.Sc. Part – I Annual Examination – 2020

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

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

Q#1 (a)	What is Design Effect (Deff)? Explain briefly.	10
(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 $\left(\frac{\lambda S}{C_1\sqrt{2\pi}}\right)^{2/3}$	10
Q#2(a)	Discuss the following types of allocations under stratified random sampling: i. Proportional Allocation ii. Optimum Allocation	10
(b)	A sampler proposes to take a stratified random sample. He expects that his field costs will be of the form $\sum c_h n_h$. His advance estimates of relevant quantities for the two strata are as follows. Stratum / W_h / S_h / C_h 1 / 0.4 / 10 / \$4 2 / 0.6 / 20 / \$9 a. Find the values of $\frac{n_1}{n}$ and $\frac{n_2}{n}$ that minimize the total field cost for a given value of $V(\bar{y}_{st})$. b. Find the sample size required, under the optimum allocation, to make $V(\bar{y}_{st}) = 1$. Ignore the fpc. c. How much will the total field cost be?	10
Q#3 (a)	What is stratified random sampling? Explain by the help of an example.	10
(b)	Given the results of a stratified random sample, an unbiased estimator of V_{ran} which is the variance of the mean of a SRS from the same population is; $v_{ran} = \frac{N-n}{n(N-1)}\left[\frac{1}{N}\sum_{h=1}^{L}\frac{N_h}{n_h}\sum_{j=1}^{n_h} y_{hj}^2 - \bar{y}_{st}^2 + v_{\bar{y}_{st}}\right]$ Where $v_{\bar{y}_{st}}$ is the usual unbiased estimator of $V_{\bar{y}_{st}}$ i.e. $E(v_{\bar{y}_{st}}) = V_{\bar{y}_{st}}$	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 $V_{(\bar{y}_{sy})} = \left(\frac{N-1}{N}\right)S^2 - \frac{k(n-1)}{N}S_{wsy}^2$ where $S_{wsy}^2 = \frac{1}{k(n-1)}\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2$	10

Q#5(a)	Discuss three methods of selecting a sample that lead to an unbiased ratio type estimate.	12
(b)	Using the approach of Hartley and Ross, obtain an upper bound for the ratio of bias in $\hat{R}_c$ to its standard error.	08
Q#6 (a)	Show that the estimator $\bar{y}_{lr}$ has a bias of order $\frac{1}{n}$ in simple random sampling.	10
(b)	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 $$V(\bar{y}_{lr}) = \frac{1-f}{n}(S_y^2 - 2b_0S_{yx} + b_0^2S_x^2)$$	10
Q#7 (a)	What is non-response error? Also describe the sources of non-response.	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)	Describe a practical example in which two-stage sampling can be used.	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. Inverse Sampling ii. Determination of Sample Size iii. Double sampling iv. Method of controlled selection	5 each