[61/A-24] SEAT No._____ No. of Printed Pages : 3

SARDAR PATEL UNIVERSITY
B.Sc. Semester – IV Examination
Saturday, 15th April, 2017
US04ESTA04
(Biostatistics – II)

Time: 2.00 to 4.00 p.m M.Marks: 70

Note: (i) Statistical table will be allowed /provided on request (ii) Simple/Scientific calculator is allowed.
(iii) Q.3 to 6 each sub question have 5 marks.

Q.1 Multiple Choice Questions (10×1)

(1) In hypothesis testing, the hypothesis tentatively assumed to be true is
(a) the alternative hypothesis (b) the null hypothesis
(c) either the null or the alternative (d) None of these alternatives is correct.

(2) In the regression $Y = 4X + 2$, what does the 2 represent?
(a) Y intercept (b) Slope of the line
(c) Any value of the independent variable that is selected (d) None of the above

(3) A group of 10 men were given a special diet for two weeks to test weight loss in pounds. The observed data was:

Man	1	2	3	4	5	6	7	8	9	10
Weight before	181	171	190	187	210	202	166	173	183	184
Weight after	178	172	185	184	201	201	160	168	180	179

To determine if the data provide sufficient evidence to indicate the special diet leads to a weight loss, the appropriate test procedure is :
(a) unpaired t - test (b) paired t - test
(c) both (a) and (b) (d) None of these

(4) The area under the normal curve between $z = -1$ and $z = 0$ is _____ the area under the normal curve between $z = 0$ and $z = 1$.
(a) Less than (b) Greater than (c) Equal to (d) None of these

(5) If $r_{XY} = -0.84$ then $r_{YX} = ?$
(a) -0.84 (b) 0.84 (c) 0.48 (d) None

(6) Under what conditions would you use the paired t-test?
(a) When there is a single sample of data (b) When the two samples of data are independent
(c) When there are two proportions (d) when the two samples of data are not independent

(7) The regression lines of X on Y and Y on X
(a) Do not intersect (b) Intersect at any point
(c) Intersect at $(\overline{X}, \overline{Y})$ (d) None of these

(8) If the random variable Z is the standard normal score, which of the following probabilities could easily be determined without referring to a table?
(a) $P(Z > 2.86)$ (b) $P(Z < 0)$ (c) $P(Z < -1.82)$ (d) $P(Z > -0.5)$

(9) The degrees of freedom in t - test for testing specified mean for single population is
(a) $n + 1$ (b) $(n - 1)/2$ (c) $n - 1$ (d) $2n - 1$

(10) Which of the following values cannot occur in a chi square distribution?
(a) 38.4 (b) 0.61 (c) 110 (d) -2.45

Q.2 Short Type Questions (Attempt Any Ten) (10×2)

(1) What is Scatter plot (diagram)? Write down its limitations.

(2) Two types of drugs were used in 5 patients each for reducing their weights. The decrease in the weight after using the drugs for six months was recorded as given below:

Drug - A	11	13	12	14	10
Drug - B	12	9	8	15	9

Identify the objective of the study. Which statistical test would you use for the same? Write down the required test procedure.

(3) The coefficient of correlation between two variables X and Y is 0.72, the covariance is 36 and the variance of X is 36, then find the standard deviation of Y.

(4) State any four properties of Normal distribution.

(5) Define two types of errors in the testing of hypothesis.

(6) What does a coefficient of correlation 0.7 means?

(7) Write down the teststatistic would you use for the following:

(i) Comparing two population means based on large samples from the two populations respectively.

(ii) Comparing two population means based on small samples from the two populations respectively.

(8) List out the various properties of regression coefficients.

(9) Juhi earned a score of 940 on a national achievement test. The mean test score was 850 with a standard deviation of 100. What percentage of students had a higher score than Juhi? (Assume that test scores are normally distributed.)

(10) Let $X \sim N(86, 25)$, determine $P(70 < X < 80)$.

(11) State the conditions under which paired t – test is used.

(12) Write in brief on chi square test in a 2×2 contingency table.

Q.3(a) Define (i) correlation coefficient (ii) Regression equation of X on Y.

(b) From the following data calculate (i) the correlation coefficient and comment on it (ii) two regression coefficients (iii) predict the height of oak tree at the age of 43 years.

Oak Tree	1	2	3	4	5	6	7	8	9	10	11	12
Age (years)	97	93	88	81	75	57	52	45	28	15	12	11
Height (inches)	12.5	12.5	8.0	9.5	16.5	11.0	10.5	9.0	6.0	1.5	1.0	1.0

OR

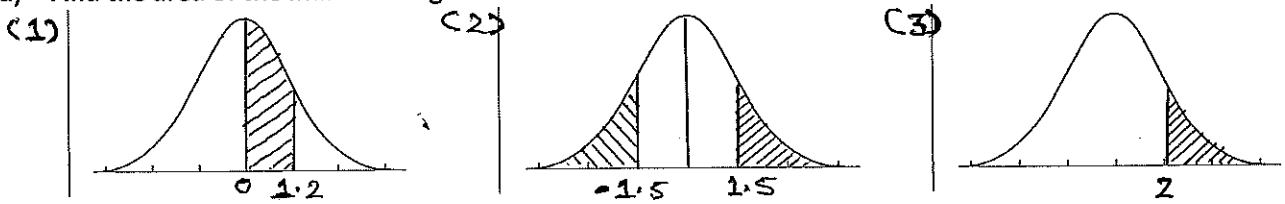
Q.3(a) Write a note on Karl Pearson's correlation coefficient method.

(b) A random sample of eight drivers insured with a company and having similar auto insurance policies was selected. The following table lists their drier experiences (in years) and monthly auto insurance premiums (in dollars)

Driving Experience	5	2	12	9	15	6	25	16
Monthly Insurance Premiums	64	87	50	71	44	56	42	60

(i) Does the insurance premium depends on driving experience? Justify your answer by calculating most suitable statistical measure (ii) find the least squares regression line by choosing appropriate dependent and independent variables (iii) Predict the monthly auto insurance premium for driver with 10 years of driving experience.

Q.4(a) Find the area of the indicated region under the standard normal curve.



(b) Studies have determined that the average gestation period (time between conception and giving birth to a child) for humans to be 266 days, with standard deviation is about 16 days. If the gestation period is normally distributed, then find the percentage of pregnancies which would be (i) between 260 to 280 days (ii) more than 39 weeks?

OR

Q.4(a) The lengths of Atlantic croaker fish are normally distributed with a mean of 10 inches and standard deviation of 2 inches. An Atlantic croaker fish is randomly selected.
Find the probability that the length of randomly selected fish is (i) less than 7 inches (ii) between 7 and 15 inches (iii) more than 15 inches.

(b) Pulse rates of adult men are normally distributed with a mean of 70 and standard deviation of 8. Find the probability that a randomly selected man have pulse rate (i) greater than 78 (ii) between 60 to 72
(iii) More than median.

Q.5(a) Write a note on unpaired t – test.

(b) A study was performed on 200 school students to investigate whether regular vitamin – A supplementation was effective in preventing colds. 100 were randomized to receive daily vitamin – A supplements during the

month of march and other receive placebo (and did not receive vitamin – A). The no. of students getting at least one cold was computed in the two groups, and the results are given in the following 2×2 contingency table.

	Cold	No cold	Total
Vitamin - A	15	85	100
Placebo	25	75	100
Total	40	160	200

Is vitamin – A supplement is effective in preventing colds? Test at $\propto = 0.05$

OR

Q.5 Recorded from a maternity hospital for 91 births in a given week indicated the following.

Mother's age (years)	Type of birth	
	No complications	With complications
Under 25	42	3
25 - 29	14	1
30 - 34	8	4
35 & over	12	7

(i) Test at 5% level of significance, the hypothesis that the incidences of complications are independent of the age of the mother (ii) If the null hypothesis is rejected after an analysis of the data, can we conclude that the incidence of complications is higher in upper age-group (35 & over) than in lower age-group (Under 25)?

Q.6(a) A sample of 8 patients had their lung capacity measured before and after a certain treatment with the following results:

Patient	1	2	3	4	5	6	7	8
Before	750	860	950	830	750	680	720	810
After	850	880	930	860	800	740	760	800

Use t - test to test the hypothesis that the treatment provides no increase in lung capacity.

(b) A medical researcher wishes to see whether the pulse rates of smokers are higher than the non smokers. Samples of 100 smokers and 100 non smokers are selected. The results are shown below:

Smokers	Non - smokers
$n_1 = 100$	$n_2 = 100$
$\overline{X_1} = 90$	$\overline{X_2} = 88$
$S_1 = 5$	$S_2 = 6$

Can the researcher conclude that smokers have higher pulse rate than non – smokers? Test at $\propto = 0.05$

OR

Q.6(a) The albumin blood levels of 10 dialysis patients are:

39	36	34	30	28	33	34	29	21	32

Perform a one – sided test of the hypothesis that the mean albumin level among the dialysis patients is greater than or equal to 35?

(b) Two diets were to be compared. Seventy five individuals were selected at random from a population of overweight people. Forty of this group were assigned to diet A and the remaining thirty five were placed on diet B. The weight losses in pounds over a period of one week were found and the following information was recorded.

	Sample size	Sample mean(lbs)	Sample variance
Diet - A	40	10.3	7.00
Diet - B	35	7.3	3.25

Is Diet – A is better than diet – B? Test at $\alpha = 0.05$.

———— X — X ————