[A-14] SEAT No._____ SC

No. of Printed Pages : 03

Sardar Patel University

B.Sc. Semester – IV Examination (NC - 2010 Batch)

Time: 10.00 to 1.00 P.M | Thursday , 11th April, 2019 | Maximum Marks: 70

Course code: US04CSTA01

(Statistical Techniques)

Note: (i) Graph paper will be provided on request (ii) Q.3 to Q.8 each sub question is of 4 marks.

Q.1 Multiple Choice Questions (10×1)

(1) The two lines of regression are $X + 2Y = 5$ and $2X + 3Y = 8$ then the correlation coefficient between X and Y is
(a) 0.75 (b) -0.75 (c) 0.87 (d) -0.87

(2) Which of the following distribution used in the construction of charts for number of defects per item is
(a) Binomial (b) Poisson (c) Normal (d) All of these

(3) Time series analysis is used to
(a) understand the past behaviors of time series (b) understand the present situation
(c) predicting the future values of the series (d) All of the above

(4) If all the points in a scatter diagram lie on the least squares regression line, then the coefficient of correlation must be
(a) 1 (b) -1 (c) 0 (d) either -1 or 1

(5) Which type of control chart should be used when it is possible to have more than one mistake per item
(a) p – chart (b) np – chart (c) $\bar{X}$ – chart (d) C – chart

(6) In the regression line $Y = a + bX$, the following is always true:
(a) $\sum(Y - \hat{Y}) = 0$ (b) $\sum Y = \sum \hat{Y}$ (c) $\sum(Y - \hat{Y})^2$ is minimum (d) All of the above

(7) With reference to SQC, the probability that any seven consecutive points lie on either side of central line is
(a) $\frac{1}{2}$ (b) $\frac{1}{2^7}$ (c) $\frac{1}{2^6}$ (d) none

(8) In semi – average method, if the time series data contains odd number of values then we drop
(a) First Value (b) Last Value (c) Middle Two Values (d) Middle Value

(9) Variation due to assignable causes in the product occurs due to
(a) Faulty process (b) Poor quality of raw materials (c) Carelessness of operators (d) All of the above

(10) The ranks given by two different judges to five participants in a debate contest are:

R1	3	1	4	2	5
R2	3	5	2	4	1

The rank correlation coefficient between them is
(a) 0 (b) 1 (c) -1 (d) Can't possible

Q.2 Short Type Questions (Attempt Any Six) (6×2)

(a) What does a coefficient of determination 0.81 means?

(b) Fifteen samples of size three are selected from a production line.
(a) What is the value of the A_2 factor used to determine lower and upper control limits for mean?
(b) What are the values of D_3 and D_4 factors used to determine the lower and upper control limits for range?

(c) The least squares trend line for an annual time series data regarding sale of cars (In' 000 units) from 2010 to 2017 is $Y = 1.2 + 0.5X$. Identify an intercept and slope of this trend line. Interpret the slope in this trend line.

(d) What do you mean by Time Series analysis?

(e) If one or more points fall below LCL in construction of p chart, what would you conclude from that?

(f) What are the chief sources of assignable causes of variations?

(g) List out the uses of SQC.

(h) How will you estimate the unknown constants in the equation $Y = f(x)$? Write in brief about it?

Q.3(a) What is curve fitting? Obtain normal equations for fitting of straight line.

(b) The following table shows the number of diabetic patients ('000) in a city from 2003 to 2012. Fit an equation of the form $Y = aX^b$ and forecast the number of diabetic patients for the year 2013 and 2014.

Year (X)	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
No. of patients (Y)	4.5	6.1	8.2	11.0	14.9	20.1	27.1	36.6	49.4	66.7

OR

Q.3(a) Obtain normal equations for estimating unknown constants in the equation $Y = a + bX + cX^2$

(b) The following table shows literacy rate of India from 1921 to 1961.

Census year (X)	1921	1931	1941	1951	1961
Literacy Rate (Y)	11.4	12.1	13.9	17.3	18.0

Fit an equation of straight line to the data given using least squares method and estimate the literacy rate for the census year 1971 and 1981. Also obtain average annual increase in the literacy rate.

Q.4(a) In usual notation, Prove that $\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$, Interpret the cases when $\rho = -1, 1$

(b) If X and Y are two independent variables with variances 36 and 16 respectively. Calculate the coefficient of correlation between U and V where $U = 2X - 3Y$ and $V = 2X + 3Y$.

OR

Q.4(a) What is correlation coefficient? Prove that correlation coefficient is an independent of change of origin and change of scale.

(b) The following table consists of one student athlete's time (in minutes) to swim 2000 yards and the student's heart rate (beats per minute) after swimming on a random sample of 10 days.

Swim time	34.12	35.72	34.72	34.05	34.13	35.73	36.17	35.57	35.37	35.57
Heart rate	144	152	124	140	152	146	128	136	144	148

(i) Identify an independent and dependent variable (ii) Does there appear to be any evidence of linear relationship between these two variables? Justify your answer by calculating most suitable statistical measure.

Q.5(a) Do as directed:

(i) Prove that if one of the regression coefficient is greater than one than the other one must be less than one.

(ii) The tangent of an angle between two regression line is given to be 0.6 and $S_Y = 2S_X$ then find the correlation coefficient between X and Y.

(b) An economist wanted to analyze the relationship between the speed of a car and its mileage. An experiment was carried out at different speeds (KMPH) and mileage (KMPL) was recorded.

Speed (X)	25	35	45	50	60	65	70
Mileage (Y)	40	39	37	33	30	27	25

* KMPH – Kilometer per hour and KMPL – Kilometer per liter

(i) Identify the objective(s) of the experiment and accordingly select an appropriate statistical measure and calculate it. (iii) find the regression equation which could be used to predict the mileage of a car when the speed is 55 KMPH? (iv) List out the variables which may influence the mileage of a car.

OR

Q.5(a) Obtain an angle between two regression lines. Interpret the cases when $r = 0, \pm 1$.

(b) The success of a shopping center can be represented as a function of the distance (in kms) from the center of the population and the number of clients (in hundreds of people) who will visit. The data given in the table below:

Distance	15	19	25	27	34	40
No. of customers	8	7	6	4	2	1

(i) Identify an independent and dependent variable (ii) compute r, the correlation coefficient

(iii) Determine the role of independent variable in the relationship that exists between these two variables. List out the other variables which can influence the dependent variable (iv) To receive 500 customers, at what distance from the center of the population should the shopping center be located?

Q.6(a) State the various causes of variations in the production process. Write in brief about them.

(b) The producer of a candy bar, reports on the package that the calorie content is 420 per 50 – gram bar. A sample of 5 bars from each of the last 10 days is sent for a chemical analysis of the calorie content. The results are shown below. Does it appear that there are any days where the calorie content is out of control? Develop an appropriate control chart and analyze your findings.

Sample	Calorie count					Sample	Calorie count				
	1	2	3	4	5		1	2	3	4	5
1	426	406	418	431	432	6	427	417	408	418	422
2	421	422	415	412	411	7	422	417	426	435	426
3	425	420	406	409	414	8	419	417	412	415	417
4	424	419	402	400	417	9	417	432	417	416	422
5	421	408	423	410	421	10	420	422	421	415	422

OR

Q.6(a) Differentiate between (i) p and np chart (ii) Variable and Attribute chart.

(b) A process produces rubber belts in lots of size 2500. Inspection records on the 16 lots reveal the data in the following table.

Lot number	No. of nonconforming Belts	Lot number	No. of nonconforming Belts
1	230	9	456
2	435	10	394
3	221	11	285
4	346	12	331
5	230	13	198
6	327	14	414
7	285	15	131
8	311	16	269

Based on the data in the table above if np chart is to be established, what would you recommend as the central line and control limits? Is the process in control?

Q.7(a) Which components of a time-series would you mainly associate each of the following? Why?

(i) A strike in a factory delaying production for five days at Manesar plant of Maruti Udyog in Haryana.

(ii) Price hike in Gold during Pusyanakshtra.

(iii) Fall in infant mortality rate due to advances in Science.

(iv) An increase in the income of government employees

(b) From the following data, find the trend values by the moving average method considering 3 yearly cycles. Write down the limitation of this method.

Year	2008	2009	2010	2011	2012	2013	2014	2015	2016
Import (in Crores)	170	231	261	267	278	302	299	298	340

OR

Q.7(a) Write a note on semi – average method.

(b) The following table shows the sales for ice – creams in three different seasons for the years 2011 through 2015. The sales are reported in millions of dollars. Calculate Seasonal trend values using least squares method.

Sales of ice – creams in three different seasons (In millions dollar)

Year	Winter	Spring	Summer
2011	6.7	4.6	10.0
2012	6.5	4.6	9.8
2013	6.9	5.0	10.4
2014	7.0	5.5	10.8
2015	7.6	6.1	11.7

Q.8(a) What do you understand by seasonal variation in time series? Explain the method of 'ratio to trend' to estimate seasonal variation in time series.

(b) From the following data, find the trend values by the method of semi - averages. Also, estimate the Facebook users for 2019. Also find average monthly increase in the users.

Number of Facebook users in India from 2011 to 2018 (In Millions)

Year	2011	2012	2013	2014	2015	2016	2017	2018
Facebook Users	135.60	165.57	194.11	219.94	242.53	261.83	277.95	291.50

OR

Q.8 Calculate the quarterly seasonal indices using ratio to trend method.

Quarter	Production of Cars (In 000 units)				
	2014	2015	2016	2017	2018
I	45	48	49	52	60
II	54	56	63	65	70
III	72	63	70	75	83
IV	60	56	65	72	86

— x —

(3)