Set 1

Unique Paper Code/Subject Code: 12273303
Name of the Paper: Data Analysis
Name of course: SEC
Scheme/Mode of Examinations: CBCS Semester – III (Admission of 2019)

Attempt total four Questions taking two from each section, A and B

Section A

1. A local hospital has released the following data on its total 314 COVID-19 patients, admitted with major complaints like Fever, Fatigue and Dry cough, along with the data on whether they were admitted to ICU (Intensive Care Unit) or Non-ICU unit as follows:

	Fever	Fatigue	Dry Cough
ICU	36	29	21
NON-ICU	100	67	61

a) Construct a contingency table using the above information
b) What symptom or complaint is more compelling to admit the patients to ICU beds?
c) If a patient has dry cough what is the possibility that he/she would require ICU admission.

(6, 5,5)

StudentSuvidha.com
downloaded from

1. एक ्सथानीय अ्सपताल ने अपने क ुल 314 COVID-19 रोगियं ो पर नन्मनललखित आंकडेजारी ककए हैं जजनमें ि ुर, थकान और स ू ि ाांसी जैसी बडी लिकायतं ो के साथ-साथ यह भी बताया िया है) कक वे िहन गिककतसा इकाई ICU) में भती ह ुये या सामा्नय इकाई में :

	Fever	Fatigue	Dry Cough
ICU	36	29	21
NON-ICU	100	67	61

क) उपरो्कत जानकारी केआधार पर एक आकज्समक तालल्का बनाइये ।
ि) कौनसे ल्षकण या लिकायत पर मरीज को ICU में ्सथानांतररत करने की आव्शयकता अगधकतम है ।
ि) यदि ककसी मरीज को स ू ि ाांसी है , तो ्कया संभावना है कक उसे ICU में ्सथानांतररत करने की जुररत ह ोगी ।

Download all NOTES and PAPERS at StudentSuvidha.com

2. a) What is trimmed mean and its advantages? How will you find the 20 % trimmed mean of {8, 93, 74, 16, 32, 39, 87, 12, 47, 50} in R.

क) छंटनी से ्कया मतलब है और इसके क्या लाभ है ? आप R । मं {8, 93, 74, 16, 32, 39, 87, 12, 47, 50} का 20% छं) टनी का माध्य mean) कैसे ननकालंि ।

b) If a constant '*c*' is added to each Y_i in a sample, yielding $Z_i = Y_i + c$, how do the sample mean and median of the Z_is relate to the mean and median of the Y_is? Verify your conjectures.

ि) यदि एक नम ने के प्रतये क Yi मंएक ज्सथरां क c जोडा जाता है ,)तो फलन yield) Zi = Yi + c बनता है । और कै से Zis का माध्य और माज्धयका Yis के माध्य और माज्धयका से संबंगधत होता है ? अपने अन मुान को लस्दध कीजजये ।

c) What is the difference between the Stem-and -Leaf display and Histogram

(िStem-and -Leaf प्रिश) न और दह्सटोग्राम Histogram) केबिी अं तर ्कया है

(10,3,3)

3. a) Suppose the Dean of the Students at D-university mailed a survey to a total of 400 students. The sample included 100 students randomly selected from each of the UG-1, UG-2, UG-3, PG-1 and PG-2 classes on campus.

i. What type of sampling was used? How would you carry out the sampling according to the method involved there?

ii. Explain why the sampling method stated in (i) is the most efficient method.

क) मान लीजजए ककD - ववश्‍वववद्यालय मं) छार्त अगध्ठाता Dean of Students) ने 400 छार्तं ो को एक सवक्षेण का मेल ककया। पंनतियन मं UG-1, UG-2, UG-3, PG-1 और PG-2कक्षाओं मं से 100)छार्तं ो को यादृजछछक random) तरीकेसे ि ना ुिया ।

(i))प्रनतियन sampling) मंककस ववगध उपयोि ककया िया था ? (i) बताए िए तरीके के अन)ुसार आप ककस तरह से पंनतियन sampling) करंि ?

(ii)) बताइए कक) मं)बताई िई पंनतियन sampling) ववगध सबसे कारिर ववगध ्कयंो है

b) Explain the benefit of 'file.choose()' with example in R?

(ि R मंउिाहरण के साथ 'File.choose ()' का लाभ बताएं ?

c) What is the use of sep = ',' and what = 'char' commands in R.

(िR मंsep = ',' और what ='char' कमांड का ्कया उपयोि है ?

(10,3,3)

Downloaded from StudentSuvidha.com

Download all NOTES and PAPERS at StudentSuvidha.com

Section B

4. The merchandise trade as a share to GDP (in %), in the eight South Asian Countries in 2018, extracted from the World Development Indicators data bank, reports the following:

Country Name	Merchandise trade (% of GDP) in 2018	
Afghanistan	42.6	Mean= 43.9
Bangladesh	36.4	Median= 40.6
Bhutan	67.6	Standard deviation= 14.4
India	30.9	
Maldives	61.9	
Nepal	46.3	
Pakistan	26.5	
Sri Lanka	38.6	

a) Compute coefficient of variation, Q1, Q3 and Z scores.
b) Are there any outliers? Explain.
c) Are the data skewed? If so, how?

4 सकल घरे ल)ू उतपाि GDP) मंदेहसिारी (% मं के (रूप मं उतपािं ो के व्यापार मं आठ िष्कष्कण एलियाई िं ो मं 2018 मं वव्शव ववकास सिसूकांक डेटा बंक से ननकाले िए आकं डे नन्मनानसुार दिए िए हं:

ि/नाम	उतपािं ो का ्वयापर (% of GDP) in 2018	
Afghanistan	42.6	मा्धय (Mean)= 43.9
Bangladesh	36.4	माज्धयका(Median)= 40.6
Bhutan	67.6	Standard deviation= 14.4
India	30.9	
Maldives	61.9	
Nepal	46.3	
Pakistan	26.5	
Sri Lanka	38.6	

(क Q1, Q3 और Z स्कोर के ि्णुांक के पररवतश्न coefficient of variation) की िणना कीजजये।
) ि) ्क या यहाँ ा कोई ौििर outlier) है ? समझाइये।
ि) ्क या आंकड़े ि्म (Skewed) है ? यदि हां , तो कै से? (10, 3, 3.5)

downloaded from StudentSuvidha.com

Download all NOTES and PAPERS at StudentSuvidha.com

5. a) A Protein Supplement producing company wants to estimate the mean amount of time that consumers spend on exercising daily. From past studies, the standard deviation is estimated as 30 minutes.
i. What sample size is needed if the company wants to be 95% confident of being correct to within ±5 minutes?
ii. If 99% confidence is desired, how many more or less consumers need to be selected than found in (a)?
b) What is the use of as.numeric(), as.integer() and history () commands in R. Give example of each. (5,5.5,6)

एक प्रोटीन अनुपूरक उत्पादक कंपनी उस समय की अनुमानित राशि का अनुमान लगाना चाहती है जो उपभोक्ता रोजाना व्यायाम करने में खर्च करते हैं। पिछले अध्ययनों से, मानक विचलन standard Deviation) 30 मिनट के रूप में अनुमानित है|

क) यदि कंपनी ±5 मिनट के भीतर सही होने के लिए 95% आश्वस्त होना चाहती है तो कितनी संख्या प्रतिदर्श की आवश्यकता है ?

ख) यदि 99% विश्वास वांछित है तो (अ) से कितने अधिक या कम उपभोक्ताओं का चयन किया जाना चाहिए

(ब) R में as.numeric(), as.integer() and history () कमांड का क्या उपयोग है ? प्रत्येक को उदाहरण सहित समझाइये |

6. A frequent flyer investigated round-trip fares for nonstop travel from New Delhi to 10 different Indian destinations on Air India and Indigo airlines.

t-Test: Paired Two Sample for Means		
	Air India	*Indigo*
Mean	886	1043.2
Variance	31918.22222	247715.7333
Observations	10	10
Pearson Correlation	0.906494641	
Hypothesized Mean Difference	0	
df	9	
t Stat	-1.444549248	
P(T<=t) one-tail	0.091241084	
t Critical one-tail	1.833112933	
P(T<=t) two-tail	0.182482167	
t Critical two-tail	2.262157163	

a) Write only the null and alternative hypothesis to test mean difference greater than zero from the given data.
b) Display regions of rejection and non-rejection that you have solved in part B.
c) At the 0.05 level of significance, is there evidence of a difference in the mean round-trip fare between Air India and Indigo airlines? (3, 3.5,10)

downloaded from StudentSuvidha.com

Download all NOTES and PAPERS at StudentSuvidha.com

6. एयर इं डडया और इंडडिो एयरलाइंस के 10 अलि-अलि भारतीय िंतव्यंो के ललए नई दिल्ली से नॉन्सटॉप यार्ता के ललए एक लिातार उडने वाले ककराए की जांि की िई।

t-Test: Paired Two Sample for Means		
	Air India	*Indigo*
Mean	886	1043.2
Variance	31918.22222	247715.7333
Observations	10	10
Pearson Correlation	0.906494641	
Hypothesized Mean Difference	0	
df	9	
t Stat	-1.444549248	
P(T<=t) one-tail	0.091241084	
t Critical one-tail	1.833112933	
P(T<=t) two-tail	0.182482167	
t Critical two-tail	2.262157163	

क) दिए िए आंकडंो मंे न्यू से अगधक अं तर का परीक्षण करने के ललए िून्य और) वैकज्लपक पररक्लपना (null and alternative hypothesis) बताइये ।

ि) अस्वीकृत और िैर-अस्वीकृत के क्षेर्त िाशइए जो आपने भाि B (part -B) में हल ककए हैं।

(ि) 0.05 के स्तर पर , क्या एयर इं डडया और इंडडिो एयरलाइंस के बीि वापसी यार्ता के अं तर का साक्ष्य है?

downloaded from StudentSuvidha.com