

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)

1. एक स्थानीय अस्पताल ने अपने कुल 314 COVID-19 रोगियों पर निम्नलिखित आंकड़े जारी किए हैं, जिनमें बुखार, थकान और सूखी खांसी जैसी बड़ी शिकायतों के साथ-साथ यह भी बताया गया है कि वे गहन चिकित्सा इकाई (ICU) में भर्ती हुये या सामान्य इकाई में :

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

- क) उपरोक्त जानकारी के आधार पर एक आकस्मिक तालिका बनाइये ।
ख) कौनसे लक्षण या शिकायत पर मरीज को ICU में स्थानांतरित करने की आवश्यकता अधिकतम है ।
ग) यदि किसी मरीज को सूखी खांसी है, तो क्या संभावना है कि उसे ICU में स्थानांतरित करने की जरूरत होगी ।

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 क्या मतलब है और इसके क्या लाभ हैं ? आप R | में {8, 93, 74, 16, 32, 39, 87, 12, 47, 50} का 20% छंटनी का माध्य (mean) कैसे निकालेंगे ।

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_i s relate to the mean and median of the Y_i s? Verify your conjectures.

ख) यदि एक नमूने के प्रत्येक Y_i में एक स्थिरांक c जोड़ा जाता है, तो फलन(yield) $Z_i = Y_i + c$ बनता है | और कैसे Z_i s का माध्य और माध्यिका Y_i s के माध्य और माध्यिका से संबंधित होता है? अपने अनुमान को सिद्ध कीजिये ।

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

ग) Stem-and -Leaf प्रदर्शन और हिस्टोग्राम (Histogram) के बीच अंतर क्या 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 लीजिए कि D - विश्वविद्यालय में छात्र अधिष्ठाता (Dean of Students) ने 400 छात्रों को एक सर्वेक्षण का मेल किया। प्रतिचयन में UG-1, UG-2, UG-3, PG-1 और PG-2 कक्षाओं में से 100 छात्रों को यादृच्छिक(random) तरीके से चुना गया ।

(i) प्रतिचयन(sampling) में किस विधि उपयोग किया गया था? (i) बताए गए तरीके के अनुसार आप किस तरह से प्रतिचयन (sampling) करेंगे ?

(ii) बताइए कि (i) में बताई गई प्रतिचयन(sampling) विधि सबसे कारगर विधि क्यों है।

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

ख) R में उदाहरण के साथ 'File.choose ()' का लाभ बताएं?

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

ग) R में sep = ',' और what ='char' कमांड का क्या उपयोग है ?

(10,3,3)

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	

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

4 सकल घरेलू उत्पाद (GDP) में हिस्सेदारी (% में) के रूप में उत्पादों के व्यापार में आठ दक्षिण एशियाई देशों में 2018 में, विश्व विकास सूचकांक डेटा बैंक से निकाले गए आंकड़े निम्नानुसार दिए गए हैं:

देश/नाम	उत्पादों का व्यापार (% of GDP) in 2018	
Afghanistan	42.6	माध्य (Mean)= 43.9
Bangladesh	36.4	माध्यिका (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 गुणांक के परिवर्तन(coefficient of variation) की गणना कीजिये ।

ख) क्या यहाँ कोई गैर (outlier) हैं? समझाइये ।

ग) क्या आंकड़े विषम(Skewed) है? यदि हां, तो कैसे?

(10, 3, 3.5)

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 में प्रतिचयनों की आवश्यकता है?

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

ग) R में `as.numeric()`, `as.integer()` and `history ()` कमांड का क्या उपयोग है? प्रत्येक को उदहारण 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		
	<i>Air India</i>	<i>Indigo</i>
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)

6. एयर इंडिया और इंडिगो एयरलाइंस के 10 अलग-अलग भारतीय गंतव्यों के लिए नई दिल्ली से नॉनस्टॉप यात्रा के लिए एक लगातार उड़ने वाले किराए की जांच की गई।

t-Test: Paired Two Sample for Means		
	<i>Air India</i>	<i>Indigo</i>
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 गए आंकड़ों में शून्य से अधिक अंतर का परीक्षण करने के लिए शून्य और वैकल्पिक परिकल्पना (null and alternative hypothesis) बताइये ।

ख) अस्वीकृति और गैर-अस्वीकृति के क्षेत्र दर्शाइए जो आपने भाग B (part -B) में हल किए हैं।

ग) 0.05 के स्तर पर, क्या एयर इंडिया और इंडिगो एयरलाइंस के बीच वापसी यात्रा के अंतर का साक्ष्य है?