

**BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI**  
**Application of Statistics and Computers in Biology (BIO G510)**  
**FIRST SEMESTER 2016 – 2017**  
**MID-SEMESTER EXAM (CLOSED BOOK)**

**Full marks: 25      DATE: 06.10.16      DURATION: 90 Mins.**

---

- **Answer to the point**
- **Answer together all parts of same question.**

**1. Mark true or false: [1\*2=2]**

- i) If the number of observations in a data set is odd, the median cannot be accurately found out rather approximated.**
- ii) A highly peaked frequency distribution curve is known as platykurtic.**

**2. An airline knows from experience that the distribution of the number of suitcases that get lost each week on a certain route is approximately normal with mean = 15.5 and sd = 3.6.**

**What is the probability that during a given week the airline will lose less than 20 suitcases? [3]**

**3. Answer the following giving proper justification: [1\*2=2]**

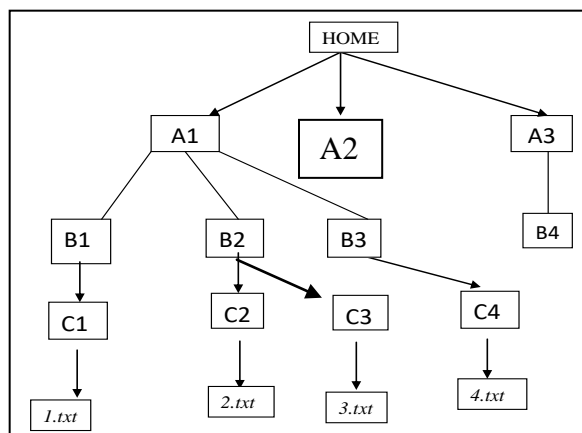
- i) In case of skewed distribution, the central tendency will be best estimated using mean or median.**
- ii) Chances of detecting significant difference is more in paired or unpaired data. Justify.**

**4. A random sample of 1,562 undergraduates enrolled in marketing courses was asked to respond on a scale from one (strongly disagree) to seven (strongly agree) to the proposition: "Advertising helps raise our standard of living." The sample mean response was 4.27 and the sample standard deviation was 1.32. Test the following hypothesis: [3]**

$H_0: \mu = 4$

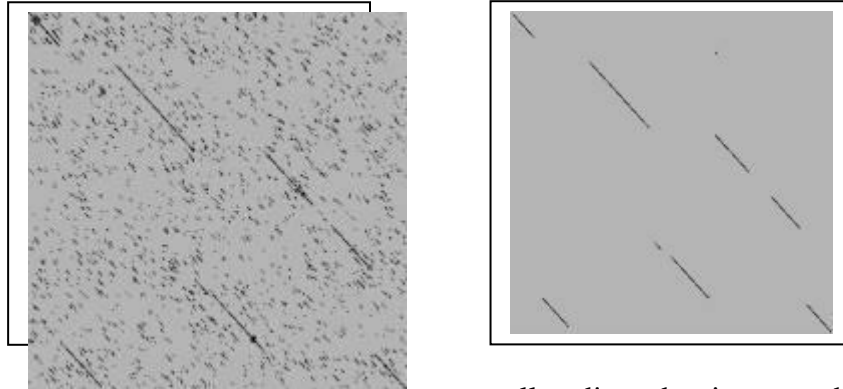
$H_A: \mu \neq 4$

**5. Consider the following file tree and answer the following questions.**



- a) Create the entire file tree in two commands. [1]  
 b) Considering that you are currently in the B3 folder, write the complete functional output of the following command. [1]  
`wc ~/A2 ~/A3 ../B1/C1/1.txt >> ../answer1.txt`

6. a) Dot matrix plots provide a quick way to visualize the similarities between two sequences. The following plots were made with a java applet with two sequences.



- i) Write two common parameters that are generally adjusted to increase the readability of the plots. Comment on the relative values of these parameters in the plots above. [1+1]  
 ii) Would you expect these sequences to have a strong (high scoring) global alignment? Justify your answer. Would a global alignment capture all significant similarities between these two sequences? [1+1]  
 b) You are shown the two alignments below. One is an alignment of two DNA sequences with an identity of 36%. The other alignment is of two amino acid sequences with an identity of 28%. Which of the two alignments represents greater biological similarity between sequences? Why? [2]

**DNA Alignment**

```
Seq1  AGGCTGCCAAAACGCACTGTTTAAT
      :  ::  :   :   ::      ::
Seq2  ACGCA-CGTTATGGCTAAAGCCTAT
```

**Amino Acid Alignment**

```
Seq3  PVALGLKEKNLYLSCVLKDKGQDIT
      :  :::           :   ::
Seq4  PADLGLMNNYNMIQLRCADLHYIT
```

c) With the help of DOT plot and DP (Needleman-Wunsch) method, establish that these two techniques reveal two different type of information. [6]

Sequence 1: AWAPQW  
 Sequence2: PPAW

[Use BLOSUM62 as a scoring matrix and -2 as gap penalty]

d) Explain each term of following expression:  $E=KN\bar{n}e^{\lambda S}$  [1]

\*\*\*\*\*

## BLOSUM62 Substitution matrix

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	4	-1	-2	-2	0	-1	-1	0	-2	-1	-1	-1	-1	-2	-1	1	0	-3	-2	0
R	-1	5	0	-2	-3	1	0	-2	0	-3	-2	2	-1	-3	-2	-1	-1	-3	-2	-3
N	-2	0	6	1	-3	0	0	0	1	-3	-3	0	-2	-3	-2	1	0	-4	-2	-3
D	-2	-2	1	6	-3	0	2	-1	-1	-3	-4	-1	-3	-3	-1	0	-1	-4	-3	-3
C	0	-3	-3	-3	9	-3	-4	-3	-3	-1	-1	-3	-1	-2	-3	-1	-1	-2	-2	-1
Q	-1	1	0	0	-3	5	2	-2	0	-3	-2	1	0	-3	-1	0	-1	-2	-1	-2
E	-1	0	0	2	-4	2	5	-2	0	-3	-3	1	-2	-3	-1	0	-1	-3	-2	-2
G	0	-2	0	-1	-3	-2	-2	6	-2	-4	-4	-2	-3	-3	-2	0	-2	-2	-3	-3
H	-2	0	1	-1	-3	0	0	-2	8	-3	-3	-1	-2	-1	-2	-1	-2	-2	2	-3
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4	2	-3	1	0	-3	-2	-1	-3	-1	3
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4	-2	2	0	-3	-2	-1	-2	-1	1
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5	-1	-3	-1	0	-1	-3	-2	-2
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5	0	-2	-1	-1	-1	-1	1
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6	-4	-2	-2	1	3	-1
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7	-1	-1	-4	-3	-2
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4	1	-3	-2	-2
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5	-2	-2	0
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11	2	-3
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	-1
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4