Indian Statistical Institute
B S D S, Second Semester, 2025-26
Mid-semester Examination
Advanced Statistical Methods I: Part I

18.03.26   **Total score 50**   **Duration: 60 minutes**

Name   **Student ID**

1. Write your name and ID on each page.

2. Numbers in brackets denote total points allotted to each question.

3. You may use calculator.

4. Laptops and phones are not allowed.

5. You are allowed to bring two pages (one A4 sheet two sided OR two A4 sheets one sided) of notes.

6. Show all your work.

Qn no	1	2	3	Total
Marks	15	15	20	50
Obtained				

1. (15) Suppose we have the following four observations and we are trying to test if they come from a uniform(0,1) distribution.

$$0.04, 0.53, 0.21, 0.66$$

Compute the Kolmogorov Smirnov test statistic showing all your calculations.

2. (2+3+5+5=15) Suppose that we have $X_i, i = 1, \cdots n$ from a uniform $(0, 1)$ distribution. We decide to use a histogram with k bins that are $1/k$ wide to estimate the density.

   (a) Let Y_j be the number of observations that fall into the j-th bin. Find the variance of Y_j .

   (b) Suppose that we have two interval counts Y_j and Y_{j+1}. Find the variance of $S = Y_j + Y_{j+1}$.

   (c) Calculate $\mathrm{Cov}(Y_j, Y_{j+1})$.

   (d) The height of the histogram is kY_j/n. Show that the variance of the height is a decreasing function of the bandwidth $1/k$.

3. (5+3+8+4=20) This dataset is drawn from a study discussed by Siegel (1956). It involves eight pairs of identical twins who are of nursery school age. In the study, for each pair, one is randomly selected to attend nursery school while the other remains at home. At the end of the study period, all 16 children are given the same social awareness test. The scores are given below. We are interested in testing the hypothesis that the median of the distribution of difference of scores between school and home trained twins is positive.

school	82	69	73	43	58	56	76	65
home	63	42	74	37	51	43	80	62

(a) Set up the problem mathematically, defining the random variables, the model and the null/alternate hypotheses.

(b) We decide to do a Wilcoxon test. What is the name of the test and what are the assumptions?

(c) Calculate the value of the test statistic.

(d) In the following code to find p-value, what should be n and T?

```
pvalue<-function(n,T){
x<-c(1:n)
k<-0
for(i in 1:n){
y<-combn(x,i)
l<-dim(y)[2]
for(j in 1:l){
k<-cbind(k,sum(x[y[,j]]))}
}
pval<-length(which(k>=T))/length(k)
return(pval)
}
```