**Assignment - I**

**(Design of Experiment)**

**Deadline: 16th Aug**

1. Generate the tables shown in class from the Titanic raw data set.

(i)Table-I: Economic Status and Sex. Compute (a) population exposed to risk, (b) Number of deaths, and (c) Deaths per 100, exposed to risk, for Male, Female, and Both.

(ii)Table-II: Economic Status and Age. Compute (a) population exposed to risk, (b) Number of deaths, and (c) Deaths per 100, exposed to risk, for Adult, Child, and Both.

Write a program in C/C++. (submit soft copies of program).

Key to Variables in titanic.dat

Column

1 Economic Status (0=crew, 1=first, 2 = second, 3 = third)

10 Age (1 = adult, 0 = child)

19 Sex (1 = male, 0 = female)

28 Survived (1 = yes, 0 = no)

Values are aligned and delimited by blanks. There are no missing values.

1. Data from the Salk vaccine field trial suggest that in 1954, the school districts in NFIP trial and in the randomized controlled experiment had similar exposures to the poliovirus.
   1. The data also show that children in the two vaccine groups (for the randomized controlled experiment and the NFIP design) came from families with similar incomes and educational backgrounds. Which two numbers in table confirm this finding?
   2. The data show that the children in the no-consent groups had similar family backgrounds. Which pair of numbers in the table confirm this finding?
   3. The data show that children in the two control groups had
   4. different family backgrounds. Which pair of numbers in the table confirm this finding?
   5. In the NFIP study, neither the control group nor the no-consent group got the vaccine. Yet the no-consent group had a lower rate of polio. Why?
   6. To show that the vaccine works, someone wants to compare the 44 / 1,00,000 in the NFIP study with the 25 / 1,00,000 in the vaccine group. What’s wrong with this idea?

# Randomized controlled double-blind experiment

**Size Rate (per 100,000)**

**Treatment** 200,000 28

**Control** 200,000 71

**No consent** 350,000 46

## The NFIP Study

### Size Rate

**Grade 2 (vaccine)** 225,000 25

**Grade 1 & 3 (control**) 725,000 54

**Grade 2 (no consent)** 125,000 44

1. From the above table, those children whose parents refused to participate in the randomized controlled Salk trial got polio at the rate of 46 / 1,00,000. On the other hand, those children whose parents consented to participation got polio at the slightly higher rate of 49 / 1,00,000 in the treatment group and control group taken together. Suppose that this field trial was repeated the following year. On the basis of figures, some parents refused to allow the children to participate in the experiment and be exposed to this higher risk of polio. Were they right? Answer yes or no, and explain briefly.
2. What is a randomized controlled double-blind experiment? How is it different from an observational study?
3. What is confounding factor? Can a confounding factor be controlled in an observational study, if yes, how?
4. Define Simpson’s paradox.