*Submit R-codes in \*.r document and Reports with interpretation in word/pdf.*

**Exercise-I Marks 40**

The “data.xls” file gives the data for a part of a large clinical trial carried out by a Radiation Therapy Oncology Group. The full study included patients with squamous carcinoma of 15 sites in the mouth and throat, with 16 participating institutions, though only data on three sites in the oropharynx reported by the six largest institutions are considered here. Patients entering the study were randomly assigned to one of the two treatment groups, radiation therapy alone or radiation therapy together with a chemotherapeutic agent. One objective of the study was to compare the two treatment policies with respect to patient survival.

**LIST OF VARIABLES:**

Variable Description

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

CASE Case Number

INST Participating Institution

SEX 1=male, 2=female

TX Treatment: 1=standard, 2=test

GRADE 1=well differentiated, 2=moderately differentiated,

3=poorly differentiated, 9=missing

AGE In years at time of diagnosis

COND Condition: 1=no disability, 2=restricted work, 3=requires

assistance with self care, 4=bed confined, 9=missing

SITE 1=faucial arch, 2=tonsillar fossa, 3=posterior pillar,

4=pharyngeal tongue, 5=posterior wall

T\_STAGE 1=primary tumor measuring 2 cm or less in largest diameter,

2=primary tumor measuring 2 cm to 4 cm in largest diameter with

minimal infiltration in depth, 3=primary tumor measuring more

than 4 cm, 4=massive invasive tumor

N\_STAGE 0=no clinical evidence of node metastases, 1=single positive

node 3 cm or less in diameter, not fixed, 2=single positive

node more than 3 cm in diameter, not fixed, 3=multiple

positive nodes or fixed positive nodes

ENTRY\_DT Date of study entry: Day of year and year, dddyy

STATUS 0=censored, 1=dead

TIME Survival time in days from day of diagnosis

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Code all qualitative variables as per the code definition given in above table in R.
2. Test whether survival time (TIME) varies across treatments (TX) and report descriptive statistics with p-value [Hint: t test].
3. Regress Survival time on Sex, Treatment, Age and T-Stage and report the results with interpretation. One can use table or graph to report the results [linear regression].

**Exercise-II Marks 40**

1. Using above data, test whether proportion of death (STATUS) varies across treatments (TX) and report contingency table with OR.
2. Also check whether proportion of death varies across T-Stages (T\_STAGE) and report the results by bar plot [Chi-square test].
3. Consider status as your outcome variable and check whether the odds of mortality is higher in standard treatment than test treatment adjusted for sex, Age, T-stages. One can use table or graphs to report the results [logistic regression].

**Report Marks 20**