ST662 Topics in Data Analytics
2018-19 Semester 2
Assignment Sheet 1

Due at 2pm Monday 18th February 2019

1. A dataset called Toenail.xlsx has been posted on Moodle. The dataset contains the results from an experiment comparing two methods for treating toenail dermatophyte onychomycosis (TDO). TDO is a common toenail infection that can be difficult to treat.

   There are five variables in the dataset:

   - ID: a unique identifier for each patient.
   - Time: the time point at which the response was recorded. This was either 0, 1, 2, 3, 6, 9 or 12 months.
   - Treat: The treatments were coded 1 or 0.
   - Gender: Male or Female.
   - y: the response recorded was that the infection was present (1) or not (0).

   (a) Read the data into your ST662 SAS library.

   (b) Create a variable in your dataset that is unique for each row of data. Do this by using the code:

```
data NewDatasetName;
   set OldDatasetName;
   Obs = _n_;
run;
```

   This new variable may help with the screening process.

   (c) Create SAS code to screen the data for any anomalies.

   (d) Create SAS code to deal with any anomalies that you have found. I.e. generate code to either change the observation (if there is an obvious error) or to mark the observation as missing.

Details on what you have to submit for this assignment

Submission of this assignment is in two parts:

1. Submit on Moodle the SAS programme you created to adress questions (a)-(d) above. This must be done before the start of class, i.e. BEFORE 2PM. Do not leave this until the last minute as Moodle submission will close at this time.

2. Submit a typed list of all errors you found with a brief comment on how you dealt with each. Attach a printed hard copy of your programme to the list. Submit at the beginning of the lecture at 2pm.