Lab 8: Part 2

The aim of this part of the lab exercise is to give you practical experience in Table analysis using R Studio and an R Notebook.

1 Table Analysis

The file *arm.csv* contains the results from an experiment where the direction of arm-crossing was recorded as well as the gender. This data was collected at AIMS (African Institute for Mathematical Sciences).

This data has two columns: gender and armcross direction (L or R arm on top).

Is the gender independent of the armcross direction?

2 Titanic revisited

Using (titanic-all-cols.csv). This is the data set used for the *Titanic: Machine Learning from Disaster* Competition. The same one we used last week.

The data includes the following variables:

- PassengerId
- Survived (1= survived, 0 = died)
- Pclass - Passenger's travel class
- Name
- Sex
- Age in Years
- SibSp - number of siblings / spouses aboard the Titanic
- Parch - number of parents / children aboard the Titanic
- Ticket - ticket number
- Fare - passenger fare
- Cabin - cabin number (if known)
- Embarked - the port of embarking (C = Cherbourg, Q = Queenstown, S = Southampton)

1. Load the `titanic-all.csv` data into an R notebook.

2. Explore the data numerically and graphically. Confirm the variables that are categorical and numerical/continuous and that R has read them in appropriately. (If you have not done this last week)

3. Use Table analysis on the categorical variables, for example: Survived vs Pclass or Survived vs Embarked - what would you conclude if these (Pclass or Embarked or Gender) were the **only** columns you had available?