Analysis of Titanic Dataset

A publicly available dataset of survivors of Titanic shipwreck contains information about Class, Sex, Survival status and Age. The dataset was manipulated to deal with missing values in order to perform analysis of proportion of survived based on Sex and Passenger Class.

It is not uncommon to compare probability of an outcome between 2 groups as a ratio. The odds are a ratio of probability of an even to its compliment i.e. if odds are 1 then the probability of an even is the same as its compliment, for odds greater than 1 probability is higher and lastly odds less than 1 probability is lower than the compliment. Log-odds are another way to report probability. Although, using log-odds is not intuitive at first but it has its advantages. It is difficult to understand the importance of difference between 99.9% and 99.9% but using log-odds the difference between 6.9 and 9.2 of 2.3 is much easier to see as significant rather than a small incremental change of 0.09.

The log odds relationship of survival vs. passenger class by sex is shown on Figure 1. , it is evident that females had a much greater chance of survival than males of the same class. The males of the 1st and 2nd class were about 30 % less likely to survive than female counterparts. However, chances of survival of females in 3rd class were about 20% lower than that of the other females.