

### Analysis of the Titanic Dataset

We have the dataset of passenger information of the Titanic and it includes the sex, passenger class and survival status. We want to analyze the log-odds of survival across the three passenger classes and sexes.

The odds of an event reflects the ratio of the probability of an event's occurring to the probability of its not occurring. From the output of the dataset, we can say that the probability of death to the probability of survival was 1.91 (odds of death = 1.91).

$$ODDS_{survival} = \frac{P_{survival}}{1 - P_{survival}}$$

The log-odds of an event means the natural logarithm of the probability of an event's occurring to the probability of its not occurring.

$$LOG - ODDS_{survival} = \ln \frac{P_{survival}}{1 - P_{survival}}$$

A logistic ANOVA model was fit. There was strong evidence of an interaction effect between the classes and sexes ( $p < 0.001$ ), and strong evidence of the main effects of each variable (both  $p < 0.0001$ ). And the plot of the log-odds of survival across sexes and passengers classes is presented in Figure 1. The log-odds of survival was much higher for female than male in each passenger classes. Among 1<sup>st</sup> and 2<sup>nd</sup> classes, the log-odds decreased in same rate. It means that the female were more likely to be survived than male in general and the likelihood of survival dropped proportionally from 1<sup>st</sup> class to 3<sup>rd</sup> class in both sexes.

The female log-odds decreased sharply in 3<sup>rd</sup> class. Males in 3<sup>rd</sup> class still had lower odds of survival but the odds of survival converged in the gender gap significantly. The odds of survival were low in 3<sup>rd</sup> class for both sexes.

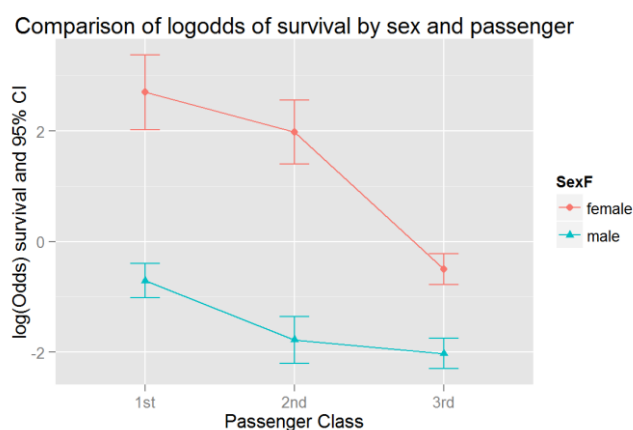


Figure 1. Comparison of log-odds of survival by sex and passenger class