

Factors Associated with Survival in the Titanic Disaster

Introduction: On the 10th of April, 1912, the RMS Titanic set out on its maiden voyage across the Atlantic Ocean, carrying 2,223 passengers. On the 14th of April, it hit an iceberg and sank. In this analysis, we examine factors associated with the probability of survival in the Titanic disaster, including age, sex and ticket class using data available on 1309 of the passengers.

Statistical Methods: Descriptive statistics were examined for each variable. Age was incorporated into the analysis as a binary variable (<15 years old vs. older). Chi-square tests were performed for preliminary inspection of potential associations between survival status and the other factors. A binary logistic regression analysis (model shown below) was performed.

Logit $[P(Y=1)] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$, where

Y = survival status (1=survived, 0 = died)

X_1 = sex (1=male, 0=female)

X_2 = age (=1 if ≥ 15 years old, 0 if < 15 years old)

X_3 = 1 if first class, 0 else; X_4 = 1 if second class, 0 else.

β_0 = true y-intercept

β_1 = difference in log(odds) (male vs. female) adjusted for age and ticket class

β_2 = difference in log(odds) (≥ 15 yrs vs. <15 yrs) adjusted for sex and ticket class

β_3 = difference in log(odds) (1st vs. 3rd class); β_4 = log(odds) difference (2nd vs. 3rd class); adjusted for sex and age

Results: Characteristics of the 1309 passengers are summarized in Table 1. 500 passengers (38%) survived. 843 (64%) were male, and 109 (10%) of the 1046 passengers for whom ages were available were children under the age of 15. 323 (25%) were first class passengers and 277 (21%) were in second class.

Table 1. Titanic Disaster: Passenger Characteristics (n=1309)

Variable	n (%)
Survived	500 (38%)
Male	843 (64%)
Age (≥ 15 years)	937 of 1046 (90%)
Ticket Class First	323 (25%)
Second	277 (21%)

Table 2. Adjusted Odds Ratios for Survival

Predictor	Adjusted Odds Ratios and 95% CI
Sex (male vs. female)	0.1 (0.06, 0.11)
Age (≥ 15 yrs vs. <15 yrs)	0.3 (0.2, 0.5)
Ticket Class: first vs. third	6.6 (4.5, 9.8)
first vs. second	2.7 (1.8, 4.0)
second vs. third	2.5 (1.7, 3.7)

The multivariable binary logistic regression model was significant (LR test p-value < 0.0001). The following were significantly associated with survival at the 5% significance level: Sex, ticket class and age ≥ 15 ($P < 0.0001$) (Table 2). With 95% confidence: (i) the female odds of surviving were between 8.9 and 17.2 times the male odds of surviving, adjusted for age and ticket class; (ii) Children younger than 15 had odds of surviving that were between 1.9 and 5.1 times the odds of older passengers, adjusted for sex and ticket class; and (iii) first class passengers' odds of surviving were between 4.5 and 9.8 times those of passengers in third class, and between 1.8 and 4.0 times those of passengers in second class, adjusted for sex and age. Second class passengers' odds of surviving were between 1.7 and 3.7 times the odds of passengers in third class, adjusted for age and sex.

The model fit was good (C statistic 0.84).

Conclusions: In the titanic disaster, the overall odds of survival were poor, with only 38.2% of the 1309 passengers on whom data were available surviving. The old adage 'women and children first' seemed to apply, with the odds of surviving for women being between 8.9 and 17.2 times the odds of surviving for men (95% confidence), adjusted for age category and ticket class, and with the odds for those < 15 years of age being between 1.9 and 5.1 times the odds of surviving of those 15 or older, adjusted for sex and ticket class. There were over 2200 passengers on the Titanic. Missing data are a limitation of this study, especially if data were disproportionately missing for particular ticket classes, ages or sex.

*BIOS 591P Spring 2022

*HW 8 Answer Key Code;

*file import wizard used to import the data
into temporary data set named 'titanic';

*proc contents for imported data set;

PROC CONTENTS DATA=WORK.titanic;

RUN;

*Create a formats for survived and sex;

PROC FORMAT;

```
VALUE survfmt 0 = 'Died'
               1 = 'Survived';
VALUE agefmt 0 = '< 15 Yrs Old'
              1 = '>= 15 Yrs Old';
VALUE sexfmt 0 = 'Female'
              1 = 'Male';
```

RUN;

*create binary variable for age, and associate the the formats with the variables;

DATA titanic;

SET titanic;

Passenger_age = (Age >= 15); *Create dummy variable, = 1 for age >= 15, = 0 for any other age;

if Age = . then Passenger_Age = .; *if age is missing, the passenger_age is set to missing;

FORMAT Survived survfmt.;

FORMAT Passenger_age agefmt.;

FORMAT Sex sexfmt.;

RUN;

*simple descriptive statistics for each variable;

PROC FREQ DATA=WORK.titanic;

TABLES Survived TicketClass Sex Passenger_Age;

RUN;

*bivariate analyses;

PROC FREQ DATA=WORK.titanic;

TABLES (TicketClass Sex Passenger_Age)*Survived / CHISQ ODDSRatio(CL=WALD) ;

RUN;

*Compare avg ages for 3 ticket classes (one-way ANOVA);

PROC LOGISTIC DATA=WORK.titanic PLOTS=ROC; *roc option produces ROC curve;

CLASS TicketClass (PARAM=REF REF='3');

MODEL Survived(Event='Survived') = Sex Passenger_Age TicketClass;

ODDSRatio Ticketclass / DIFF=ALL; * ORs comparing ticket classes (pairwise);

RUN;