Logistic Regression Interpretation for Titanic Data

For the Titanic data we want to know which variables are good predictors of survival and how much an effect these variables have in predicting survival.

Predictors that we consider: Sex $<$ M / F Age $<$ Adult / Child

We will consider two models

Model 1 (No interaction)

log-odds of survival

$$\log\left(\frac{p}{1-p}\right) = b_0 + b_1 \text{Female} + b_2 \text{Child}$$

Adult ♂ $\rightarrow = b_0$ (Baseline)

Adult ♀ $\rightarrow = b_0 + b_1$

Child ♂ $\rightarrow = b_0 + b_2$

Child ♀ $\rightarrow = b_0 + b_1 + b_2$

$$\log\left(\frac{p}{1-p}\right) = -1.336 + 2.294 \times F + 0.556 \times C$$

Model 2 (Interaction)

intercept

$$\log\left(\frac{p}{1-p}\right) = b_0 + b_1 \text{Female} + b_2 \text{Child} + b_3 \text{Female} \times \text{Child}$$

$\rightarrow = b_0$

$\rightarrow = b_0 + b_1$

$\rightarrow = b_0 + b_2$

$\rightarrow = b_0 + b_1 + b_2 + b_3$

$$\log\left(\frac{1}{1-p}\right) = -1.369 + 2.434 \times F + 1.181 \times C - 1.746 \times F \times C$$

We interpret parameters in terms of the odds, noting

$$\frac{p}{1-p} = e^{\log\left(\frac{p}{1-p}\right)}$$

Example

Model 1

A♂ $\frac{p}{1-p} = e^{b_0}$

C♀ $\frac{p}{1-p} = e^{b_0} \times e^{b_1} \times e^{b_2}$

Model 2

$\frac{p}{1-p} = e^{b_0}$

$\frac{p}{1-p} = e^{b_0} \times e^{b_1} \times e^{b_2} \times e^{b_3}$

(Model 1)

In general, with non interaction terms, we can describe the influence of a single variable by looking e^{b_i}

Adult Male

$$e^{-1.336} = \frac{p}{1-p} = 0.26277$$

Adult Female

$$\frac{p}{1-p} = e^{-1.336} \times e^{2.294}$$

We increase the odds by $e^{2.294} = 9.915$

Child Male

$$\frac{p}{1-p} = e^{-1.336} \times e^{0.556}$$

We increase the odds by $e^{0.556} = 1.744$

Child Female

We increase odds by $e^{0.556} \times e^{2.294} = e^{0.556 + 2.294}$

We increase the odds by $e^{2.294} = 9.915$

891.5% increase in the odds

We increase the odds by $e^{0.556} = 1.744$

74%

$= e^{0.556+2.294}$

$\approx 1600\%$

1628.8%

Note

$e^0 = 1$

If a term was 0, it would not affect odds

$e^{+} > 1$

$e^{-} < 1$

Model 2

$$\log\left(\frac{1}{1-p}\right) = -1.369 + 2.434 \times F + 1.181 \times C - 1.746 \times F \times C$$

When we have interaction terms things become more complex

Now

$e^{2.434}$ represents the increased odds of survival for an adult female compared to baseline (A♂)

$\approx 11.4 \rightarrow 1040\%$ increase

$e^{1.181}$ represents the increased odds of survival for a male child. compared to baseline.

$\approx 3.258 \rightarrow 226\%$ increase

$e^{-1.746}$ doesn't represent the increased (or decreased) odds of survival for female children but cannot be interpreted on its own.

≈ 0.174 82.6% decrease in odds of survival

$e^{2.434+1.181-1.746}$ represents the increased odds of survival for a female child compared to a male adult.

$\approx e^{1.868} = 6.48 \rightarrow 548\%$ increase in odds of survival

To calculate probabilities

model 2

$$p = \frac{e^{b_0+b_1 \times F+b_2 \times C+(b_3 \times F \times C)}}{1+e^{b_0+b_1 \times F+b_2 \times C+(b_3 \times F \times C)}} = \frac{1}{1+e^{-(b_0+b_1 \times F+b_2 \times C+(b_3 \times F \times C))}}$$

Model 1

A♂	A♀	C♂	C♀
0.208	0.723	0.314	0.820

What is the likelihood of

Model 1

0.208 0.723 0.314 0.820 ☆

Model 2

0.203 0.744 ☆ 0.453 0.622

What is the actual odds of survival for each group?

The model can fit this exactly because we have as many parameters in the model as we have value of odds to predict.

4 (A♂, A♀, C♂, C♀)

4 (b_0, b_1, b_2, b_3)