**Lab #1.2: Logistic Regression in R**

# Part 5: Try it yourself:

**Looking at the results returned by model2.summary(), and try to interpret the output.**

Table

Description automatically generated

**Questions:**

(1) What is the dependent variable and what are the independent variables?

The dependent variable is survived and the independent variables are age, pclass, male, SibSp, Parch, Fare

(2) How many observations are there in this dataset?

714

(3) What does the Intercept of 5.3890 mean?

The intercept of 5.3890 indicates that the log odds of survival when all the independent variables are equal to zero. The intercept value itself does not have any meaningful interpretation as it is an artifact of the model and the chosen reference values for the categorical variables.

(4) Which predictor variables are significantly correlated with Survived? Explain the coefficients of those variables.

The predictor variable that is significantly correlated with survived is age, Pclass, Male, and SibSp. They provide information on the direction and strength of the association.

Age: The coefficient of age is -0.0440, which means that for every unit increased in age the log odds of surviving decrease by 0.0440, while all the other constants are constant.

Pclass: The coefficient of Pclass is -1.2422, which means that for every unit increased in Pclass the log odds of surviving decrease by 1.2422, while all the other constants are constant.

Male: The coefficient of male is -2.6348, which means that for every unit increased in male the log odds of surviving decrease by 2.6348, while all the other constants are constant.

SibSp: The coefficient of SibSp is -2.6348, which means that for every unit increased in SibSp the log odds of surviving decrease by 2.6348, while all the other constants are constant.

(5) If your goal is to test the “effect” of age on people’s propensity to survive in Titanic, are there any other variables we should include in this model to avoid omitted variable bias? List 3 of them and explain your reasons.

Weight: for people with extremely high weight, it might be better to save more people with a relatively lower weight.

Income: People with higher income are more likely to be rescued

Health: People who are more likely to live a healthy life and are in general not suffering from any ailments are more likely to be rescued.

(6) After you include the above three variables, do you think this model accurately tested the “effect” of age on people’s propensity to survive in Titanic? Why?

No. The results show correlation only, as there could be variables that were omitted or neglected, even after including the three variables, but it is not controlled by the model.