

Statistics 4590: *Titanic* Survival Data

This data set gives the survival status of passengers on the Titanic, which sank off the coast of Newfoundland in 1912. You are provided with the names of the passengers, along with their ages, gender, passenger class and survival status.

You can find much more information on this data set from the following source:

Hinde, Philip (1998). *Encyclopedia Titanica*, found at www.encyclopedia-titanica.org

The datafile has the following form:

Variable	Description
-----	-----
Name	Recorded Name of Passenger
PClass	Passenger Class: 1st, 2nd or 3rd
Age	Age in years
Sex	male or female
Survived	1 = Yes, 0 = No

About half of the ages for the 3rd Class passengers are missing, although many of these could be filled in from the information of Hinde (1998). I will not expect you to do that, though.

The data is on the course webpage, and can be loaded into R using the **read.table** command.

Your goal is to see how survival status is associated with the other variables. Some possible ways to study this include:

1. Is survival associated with passenger class?
2. Is survival associated with age?
3. Is survival associated with sex?
4. Can you develop a model to examine the combined effect of all the variables on survival status? If so, what model appears most appropriate for predicting survival? Do the assumptions of this model appear to be satisfied?

When answering these questions, be sure to address the issue of whether these associations are “real”, or could other variables be confounding the results.