### **Instructions:**

=> If course materials brought you here, scroll or use the bookmarks to section.

Titanic Data	
Titaliic Data	

### M.2.Assignment - Titanic Data

M Weekly Topic & Assignment M2

## Titanic Data

The titanic and titanic2 data frames describe the survival status of individual passengers on the Titanic. The titanic data frame does not contain information from the crew, but it does contain actual ages of half of the passengers. The principal source for data about Titanic passengers is the <code>Encyclopedia Titanica</code>. The datasets used here were begun by a variety of researchers. One of the original sources is Eaton & Haas (1994) <code>Titanic: Triumph and Tragedy, Patrick Stephens Ltd, which includes a passenger list created by many researchers and edited by Michael A. Findlay.</code>

The variables on our extracted dataset are pclass, survived, name, age, embarked, home.dest, room, ticket, boat, and sex. pclass refers to passenger class (1st, 2nd, 3rd), and is a proxy for socio-economic class. Age is in years, and some infants had fractional values. The titanic2 data frame has no missing data and includes records for the crew, but age is dichotomized at adult vs. child. These data were obtained from Robert Dawson, Saint Mary's University, E-mail. The variables are pclass, age, sex, survived. These data frames are useful for demonstrating many of the functions in Hmisc as well as demonstrating binary logistic regression analysis using the Design library. For more details and references see Simonoff, Jeffrey S (1997): The "unusual episode" and a second statistics course. J Statistics Education, Vol. 5 No. 1.

Thomas Cason of UVa has greatly updated and improved the titanic data frame using the *Encyclopedia Titanica* and created a new dataset called titanic3. These datasets reflects the state of data available as of 2 August 1999. Some duplicate passengers have been dropped, many errors corrected, many missing ages filled in, and new variables created. Click <a href="here">here</a> for information about how this dataset was constructed.

An interesting result may be obtained using functions from the Hmisc library

- attach(titanic3)
- plsmo(age, survived, group=sex, datadensity=T) # or group=pclass
- plot(naclus(titanic3)) # study patterns of missing values
- summary(survived ~ age + sex + pclass + sibsp + parch, data=titanic3)

### Frank E Harrell Jr

Last modified: Fri Dec 27 16:24:38 EST 2002

### References:

Titantic data from: <a href="https://hbiostat.org/data/repo/titanic.html">https://hbiostatistics datasets</a>. Retrieved from: <a href="https://hbiostat.org/data/">https://hbiostat.org/data/</a>. Retreieved on 05.15.2023

Master templates

ny.header.2

my.header.2				
Wk	Weekly Topic & Assignment			
0				
"				

my.header.1

my.neader.i		
Wk	Focus & Medium	Weekly Topic & Assignment

# Wk Weekly Topic & Assignment Header: 1) What is 1) Perhaps one of th 2) What isn't a) Getting your n 3) The mechanics and process • Orient 4) problem Templated techniques belo you quickly

# 

https://www.utk.edu/ 3 of 3