
[\[http://biostat.mc.vanderbilt.edu/wiki/pub/Main/DataSets/titanic.html\]](http://biostat.mc.vanderbilt.edu/wiki/pub/Main/DataSets/titanic.html)

Titanic Datasets

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 Encyclopedia Titanica. The datasets used here were begun by a variety of researchers. One of the original sources is Eaton & Haas (1994) Titanic: Triumph and Tragedy, Patrick Stephens Ltd, which includes a passenger list created by many researchers and edited by Michael A. Findlay.

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 here for information about how this dataset was constructed.

DATASET specs

NAME: titanic3
TYPE: Census
SIZE: 1309 Passengers, 14 Variables

DESCRIPTIVE ABSTRACT:

The titanic3 data frame describes the survival status of individual passengers on the Titanic.

The titanic3 data frame does not contain information for the crew, but it does contain actual and estimated ages for almost 80% of the passengers.

SOURCES:

Hind, Philip. *Encyclopedia Titanica*. Online-only resource. Retrieved 01Feb2012 from <http://www.encyclopedia-titanica.org/>

VARIABLE DESCRIPTIONS

Pclass	Passenger Class (1 = 1st; 2 = 2nd; 3 = 3rd)
survival	Survival (0 = No; 1 = Yes)
name	Name
sex	Sex
age	Age
sibsp	Number of Siblings/Spouses Aboard
parch	Number of Parents/Children Aboard
ticket	Ticket Number
fare	Passenger Fare (British pound)
cabin	Cabin
embarked	Port of Embarkation (C = Cherbourg; Q = Queenstown; S = Southampton)
boat	Lifeboat
body	Body Identification Number
home.dest	Home/Destination

SPECIAL NOTES

Pclass is a proxy for socio-economic status (SES)
1st ~ Upper; 2nd ~ Middle; 3rd ~ Lower

Age is in Years; Fractional if Age less than One (1)
If the Age is estimated, it is in the form xx.5

Fare is in Pre-1970 British Pounds ()
Conversion Factors: 1 = 12s = 240d and 1s = 20d

With respect to the family relation variables (i.e. **sibsp** and **parch**) some relations were ignored. The following are the definitions used for **sibsp** and **parch**.

- Sibling:** Brother, Sister, Stepbrother, or Stepsister of Passenger Aboard Titanic
- Spouse:** Husband or Wife of Passenger Aboard Titanic (Mistresses and Fiances Ignored)
- Parent:** Mother or Father of Passenger Aboard Titanic
- Child:** Son, Daughter, Stepson, or Stepdaughter of Passenger Aboard Titanic

Other family relatives excluded from this study include cousins, nephews/nieces, aunts/uncles, and in-laws. Some children travelled only with a nanny, therefore parch=0 for them. As well, some travelled with very close friends or neighbors in a village, however, the definitions do not support such relations.

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)

SUBMITTED BY: Thomas E. Cason tcason@virginia.edu

From MLF

Prior to February 15, 1971 ("Decimal day," or "D-day"), monetary amounts in the U.K. were expressed as pounds (£), shillings (s.), and pence (d.), where £1 = 20s. = 240d. After 1970, there were 100 pennies in a pound, so one (new) penny = 2.4 old pence. All numbers should be entered in decimal rather than fractional form (for example, 1.5 rather than 1 1/2).

Lawrence H. Officer. Five Ways to Compute the Relative Value of a UK Pound Amount, 1830 to Present. MeasuringWorth.com. Retrieved 2011Sep06 from <http://www.measuringworth.com/ukcompare/>