

Homework

Titanic Dataset Summary

Xsin-Yu Huang

2025-02-23

目錄

Load Data	1
Dataset Description	2
Missing Values	2

Load Data

```
library(tidyverse)

# Load Titanic dataset
titanic <- read.csv("titanic.csv")

# Summary statistics
summary(titanic)
```

PassengerId	Survived	Pclass	Name
Min. : 1.0	Min. :0.0000	Min. :1.000	Length:891
1st Qu.:223.5	1st Qu.:0.0000	1st Qu.:2.000	Class :character
Median :446.0	Median :0.0000	Median :3.000	Mode :character
Mean :446.0	Mean :0.3838	Mean :2.309	
3rd Qu.:668.5	3rd Qu.:1.0000	3rd Qu.:3.000	
Max. :891.0	Max. :1.0000	Max. :3.000	

Sex	Age	SibSp	Parch
Length:891	Min. : 0.42	Min. :0.000	Min. :0.0000
Class :character	1st Qu.:20.12	1st Qu.:0.000	1st Qu.:0.0000
Mode :character	Median :28.00	Median :0.000	Median :0.0000
	Mean :29.70	Mean :0.523	Mean :0.3816
	3rd Qu.:38.00	3rd Qu.:1.000	3rd Qu.:0.0000
	Max. :80.00	Max. :8.000	Max. :6.0000
	NA's :177		

Ticket	Fare	Cabin	Embarked
Length:891	Min. : 0.00	Length:891	Length:891
Class :character	1st Qu.: 7.91	Class :character	Class :character
Mode :character	Median :14.45	Mode :character	Mode :character
	Mean :32.20		
	3rd Qu.:31.00		
	Max. :512.33		

Dataset Description

The Titanic dataset contains information on 891, consists of 12 columns (variables) and 891 rows (observations).

There are variables in the dataset:

1. PassengerId: Unique identifier for each passenger (Nominal variable)
2. Survived: Survival status (Binary: 0 = Did not survive, 1 = Survived)
3. Pclass: Passenger class (Categorical variable: 1 = First, 2 = Second, 3 = Third)
4. Name: Passenger name (Nominal variable)
5. Sex: Gender of the passenger (Categorical variable: male, female)
6. Age: Age of the passenger (Numeric)
7. SibSp: Number of siblings/spouses aboard (Integer)
8. Parch: Number of parents/children aboard (Integer)
9. Ticket: Ticket number (Nominal variable)
10. Fare: Fare paid for the ticket (Numeric)
11. Cabin: Cabin number (Nominal variable)
12. Embarked: embarked (Categorical variable: C, Q, S)

Missing Values

```
# Check for missing values
sapply(titanic, function(x) sum(is.na(x)|(x=="")))
```

PassengerId	Survived	Pclass	Name	Sex	Age
0	0	0	0	0	177
SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	0	0	0	687	2

Age: 177 entries are missing.

Cabin: 687 entries are missing, indicating many passengers did not have assigned cabins.

Embarked: 2 entries are missing.