

HW1

R26134034 黃纓婷

Yingting Huang

2025-02-27

目錄

Github link	1
變數介紹	1
敘述統計	1

Github link

[Github](#)

變數介紹

PassengerId : 乘客 ID

Pclass : 乘客票務艙等級，分為 1.頭等艙 2.二等艙 3.三等艙

Name : 乘客姓名

Sex : 乘客性別，分為 female、male

Age : 乘客年齡

SibSp : 乘客在船上的兄弟姐妹及配偶數量，分為 0、1、2、3、4、5、8

Parch : 乘客在船上的父母及子女數量，分為 0、1、2、3、4、5、6

Ticket : 乘客的票號

Fare : 乘客票價

Cabin : 乘客的客艙號碼

Embarked : 乘客登船的港口，分為 C.Cherbourg Q.Queenstown S.Southampton


Survived : 此乘客是否存活，分為 0.死亡 1.存活

敘述統計

```
setwd("D:/ncku2021-2024/2025_spring/stat_consult/HW1")
library(reticulate)
library(Hmisc)

data <- read.csv("titanic.csv")
# summary(data)
data[data == ""] <- NA
data$Survived <- factor(data$Survived)
data$Pclass <- factor(data$Pclass)
data$Sex <- factor(data$Sex)
data$Embarked <- factor(data$Embarked)
```

```
latex(describe(data), file="")
```

12 Variables												data 891 Observations		
PassengerId														
n	missing	distinct	Info	Mean	pMedian	Gmd	.05	.10	.25	.50	.75	.90	.95	
891	0	891	1	446	446	297.3	45.5	90.0	223.5	446.0	668.5	802.0	846.5	
lowest : 1 2 3 4 5, highest: 887 888 889 890 891														
Survived														
n	missing	distinct												
891	0	2												
Value	0	1												
Frequency	549	342												
Proportion	0.616	0.384												
Pclass														
n	missing	distinct												
891	0	3												
Value	1	2	3											
Frequency	216	184	491											
Proportion	0.242	0.207	0.551											
Name														
n	missing	distinct												
891	0	891												
lowest : Abbing, Mr. Anthony Abbott, Mr. Rossmore Edward Abbott, Mrs. Stanton (Rosa Hunt)														
highest: Yousseff, Mr. Gerious Yrois, Miss. Henriette ("Mrs Harbeck") Zabour, Miss. Hileni														
Sex														
n	missing	distinct												
891	0	2												
Value	female	male												
Frequency	314	577												
Proportion	0.352	0.648												
Age														
n	missing	distinct	Info	Mean	pMedian	Gmd	.05	.10	.25	.50	.75	.90	.95	
714	177	88	0.999	29.7	29	16.21	4.00	14.00	20.12	28.00	38.00	50.00	56.00	
lowest : 0.42 0.67 0.75 0.83 0.92, highest: 70 70.5 71 74 80														
SibSp														
n	missing	distinct	Info	Mean	pMedian	Gmd								
891	0	7	0.669	0.523	0.5	0.823								
Value	0	1	2	3	4	5	8							
Frequency	608	209	28	16	18	5	7							
Proportion	0.682	0.235	0.031	0.018	0.020	0.006	0.008							
For the frequency table, variable is rounded to the nearest 0														
Parch														
n	missing	distinct	Info	Mean	pMedian	Gmd								
891	0	7	0.556	0.3816	0	0.6259								
Value	0	1	2	3	4	5	6							
Frequency	678	118	80	5	4	5	1							
Proportion	0.761	0.132	0.090	0.006	0.004	0.006	0.001							
For the frequency table, variable is rounded to the nearest 0														
Ticket														
n	missing	distinct												
891	0	681												
lowest : 110152 110413 110465 110564 110813														
highest: W./C. 6608 W./C. 6609 W.E.P. 5734 W/C 14208 WE/P 5735														

Fare

	n	missing	distinct	Info	Mean	pMedian	Gmd	.05	.10	.25
	891	0	248	1	32.2	19.6	36.78	7.225	7.550	7.910
	.50	.75	.90	.95						
	14.454	31.000	77.958	112.079						

lowest : 0 4.0125 5 6.2375 6.4375 , highest: 227.525 247.521 262.375 263 512.329

Cabin

	n	missing	distinct
	204	687	147

lowest : A10 A14 A16 A19 A20, highest: F33 F38 F4 G6 T

Embarked

	n	missing	distinct
	889	2	3

Value	C	Q	S
Frequency	168	77	644
Proportion	0.189	0.087	0.724

在本資料集中，共有 891 位乘客，其中 38.4% 存活，死亡人數多於存活人數。

- 觀察上表，可以發現
 - Cabin 缺失率高達 77.2%，可能影響此變數對存活率的分析。
 - 艙等(Pclass)分成三種，超過一半的人(55.1%)在三等艙。
 - 性別以男性居多(64.8%)，女性較少(35.2%)。
 - 約 68.2% 的乘客沒有同行的兄弟姐妹或配偶 (SibSp=0)，76.1% 沒有同行的父母或子女 (Parch=0)。
 - 年齡分布從0歲至80歲，平均29.7歲。

綜合以上觀察，推測這些變數可能會影響存活率，比如說:可以針對不同艙等的人、不同性別的人(女性是否較容易優先被救援)、不同同行的人數(獨行旅客會不會較難獲得幫助)等等做更精準的統計分析，觀察這些變數是否會影響存活率。