

# HW1

## summary the titanic data

Ping-Chieh, Tsai

2025-02-21

### 一、讀取資料

```
library(readr)
titanic <- read_csv("C:/Users/bcai3/Desktop/titanic.csv")
```

此資料在cabin這個變數有許多缺失值

### 二、觀察資料

```
head(titanic)
```

```
# A tibble: 6 x 12
  PassengerId Survived Pclass Name      Sex      Age SibSp Parch Ticket   Fare Cabin
      <dbl>     <dbl>   <dbl> <chr>    <chr>   <dbl> <dbl> <dbl> <chr>   <dbl> <chr>
1         1         0       3 Braund~ male    22     1     0 A/5 2~   7.25 <NA>
2         2         1       1 Cuming~ fema~   38     1     0 PC 17~  71.3  C85
3         3         1       3 Heikki~ fema~   26     0     0 STON/~   7.92 <NA>
4         4         1       1 Futrel~ fema~   35     1     0 113803 53.1  C123
5         5         0       3 Allen,~ male    35     0     0 373450  8.05 <NA>
6         6         0       3 Moran,~ male    NA     0     0 330877  8.46 <NA>

# i 1 more variable: Embarked <chr>
```

```
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
```

此筆資料有五個變數是character的型態，然後在PassengerID, Pclass, SibSp, Parch這四筆資料若要建立回歸模型的話可能是dummy variable