

HW1

STATISTICAL CONSULTING

Jhih-Heng Chen

2025-02-24

經由分析鐵達尼號乘客的資料，我們可以觀察到以下幾點統計結果。本次航行共有891名乘客，其中男性約占 64.8%，女性約占 35.2%。乘客的艙等分佈顯示，三等艙(55.1%)的乘客最多，再來是一等艙 (24.2%) 與二等艙 (20.7%)。然而在年齡分佈方面，乘客年齡從不到 1 歲至 80 歲不等，但有177名乘客(19.9%)的年齡資料缺失。已知年齡的乘客中，75%的年齡在38歲以下，平均年齡約為29.7歲，顯示多數乘客為年輕族群。進一步觀察家庭同行情況，68.2%的乘客沒有同行的兄弟姐妹或配偶，而76.1%的乘客沒有同行的父母或子女，顯示許多乘客是獨自搭乘這艘船的。在票價部分，此次航行的平均票價為 32.2元，但票價分佈不均，其中 中位數為14.45元，表示票價受艙等與登船地點影響較大。最後可以觀察到乘客來自不同的登船港口，其中644人(72.4%)從S港口登船，168人(18.9%)從C港口登船，77人 (8.7%)從Q港口登船。並根據生還情況的統計，本次航行共有 342名乘客存活，生存率約為38.4%，由此可知此次船難對乘客造成極大的影響。

```
library(Hmisc)
library(reticulate)

file_path <- "C:/Users/ /Desktop/ / /titanic.csv"
titanic <- read.csv(file_path, stringsAsFactors = FALSE)
summary <- describe(titanic)
latex(summary, file = "")
```

titanic												12 Variables			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	Info	Sum	Mean												
891	0	2	0.71	342	0.3838												

Pclass

n	missing	distinct	Info	Mean	pMedian	Gmd
891	0	3	0.81	2.309	2.5	0.8631

Value	1	2	3
Frequency	216	184	491
Proportion	0.242	0.207	0.551

For the frequency table, variable is rounded to the nearest 0

Name

n	missing	distinct
891	0	891

lowest : Abbing, Mr. Anthony
highest: Yousseff, Mr. Gerious

Abbott, Mr. Rossmore Edward
Yrois, Miss. Henriette ("Mrs Harbeck")

Abbott, Mrs. Stanton (Rosa Hunt)
Zabour, Miss. Hileni

Abelson, M
Zabour, Mi

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
```
