Titanic

2022-11

Titanic

Trabajaremos con el Dataset **Titanic**

Table 1: Primeros registros de la base de datos Titanic

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171	7.2500	NA	S
2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38	1	0	PC 17599	71.2833	C85	C
3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2. 3101282	7.9250	NA	S
4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803	53.1000	C123	S
5	0	3	Allen, Mr. William Henry	male	35	0	0	373450	8.0500	NA	S
6	0	3	Moran, Mr. James	male	NA	0	0	330877	8.4583	NA	Q

Data Frame Summary

TitanicT   Dimensions: 891 x 12
Duplicates: 0

No	Variable	Stats / Values	Freqs (% of Valid)	Valid	Missing
1	PassengerId [numeric]	Mean (sd) : 446 (257.4) min < med < max: 1 < 446 < 891 IQR (CV) : 445 (0.6)	891 distinct values	891 (100.0%)	0 (0.0%)
2	Survived [numeric]	Min : 0 Mean : 0.4 Max : 1	0 : 549 (61.6%) 1 : 342 (38.4%)	891 (100.0%)	0 (0.0%)
3	Pclass [numeric]	Mean (sd) : 2.3 (0.8) min < med < max: 1 < 3 < 3 IQR (CV) : 1 (0.4)	1 : 216 (24.2%) 2 : 184 (20.7%) 3 : 491 (55.1%)	891 (100.0%)	0 (0.0%)
4	Name [character]	1. Abbing, Mr. Anthony 2. Abbott, Mr. Rossmore Edwa 3. Abbott, Mrs. Stanton (Ros 4. Abelson, Mr. Samuel 5. Abelson, Mrs. Samuel (Han [886 others]	1 (0.1%) 1 (0.1%) 1 (0.1%) 1 (0.1%) 1 (0.1%) 886 (99.4%)	891 (100.0%)	0 (0.0%)
5	Sex [character]	1. female 2. male	314 (35.2%) 577 (64.8%)	891 (100.0%)	0 (0.0%)

No	Variable	Stats / Values	Freqs (% of Valid)	Valid	Missing
6	Age [numeric]	Mean (sd) : 29.7 (14.5) min < med < max: 0.4 < 28 < 80 IQR (CV) : 17.9 (0.5)	88 distinct values	714 (80.1%)	177 (19.9%)
7	SibSp [numeric]	Mean (sd) : 0.5 (1.1) min < med < max: 0 < 0 < 8 IQR (CV) : 1 (2.1)	7 distinct values	891 (100.0%)	0 (0.0%)
8	Parch [numeric]	Mean (sd) : 0.4 (0.8) min < med < max: 0 < 0 < 6 IQR (CV) : 0 (2.1)	7 distinct values	891 (100.0%)	0 (0.0%)
9	Ticket [character]	1. 1601 2. 347082 3. CA. 2343 4. 3101295 5. 347088 [676 others]	7 (0.8%) 7 (0.8%) 7 (0.8%) 6 (0.7%) 6 (0.7%) 858 (96.3%)	891 (100.0%)	0 (0.0%)
10	Fare [numeric]	Mean (sd) : 32.2 (49.7) min < med < max: 0 < 14.5 < 512.3 IQR (CV) : 23.1 (1.5)	248 distinct values	891 (100.0%)	0 (0.0%)
11	Cabin [character]	1. B96 B98 2. C23 C25 C27 3. G6 4. C22 C26 5. D [142 others]	4 (2.0%) 4 (2.0%) 4 (2.0%) 3 (1.5%) 3 (1.5%) 186 (91.2%)	204 (22.9%)	687 (77.1%)
12	Embarked [character]	1. C 2. Q 3. S	168 (18.9%) 77 (8.7%) 644 (72.4%)	889 (99.8%)	2 (0.2%)

Data Frame Summary

TitanicT   Group: Sex = female
Dimensions: 314 x 12
Duplicates: 0

No	Variable	Stats / Values	Freqs (% of Valid)	Valid	Missing
1	PassengerId [numeric]	Mean (sd) : 431 (256.8) min < med < max: 2 < 414.5 < 889 IQR (CV) : 409.5 (0.6)	314 distinct values	314 (100.0%)	0 (0.0%)
2	Survived [numeric]	Min : 0 Mean : 0.7 Max : 1	0 : 81 (25.8%) 1 : 233 (74.2%)	314 (100.0%)	0 (0.0%)
3	Pclass [numeric]	Mean (sd) : 2.2 (0.9) min < med < max: 1 < 2 < 3 IQR (CV) : 2 (0.4)	1 : 94 (29.9%) 2 : 76 (24.2%) 3 : 144 (45.9%)	314 (100.0%)	0 (0.0%)

No	Variable	Stats / Values	Freqs (% of Valid)	Valid	Missing
4	Name [character]	1. Abbott, Mrs. Stanton (Ros 2. Abelson, Mrs. Samuel (Han 3. Ahlin, Mrs. Johan (Johann 4. Aks, Mrs. Sam (Leah Rosen 5. Allen, Miss. Elisabeth Wa [309 others]	1 (0.3%) 1 (0.3%) 1 (0.3%) 1 (0.3%) 1 (0.3%) 309 (98.4%)	314 (100.0%)	0 (0.0%)
6	Age [numeric]	Mean (sd) : 27.9 (14.1) min < med < max: 0.8 < 27 < 63 IQR (CV) : 19 (0.5)	63 distinct values	261 (83.1%)	53 (16.9%)
7	SibSp [numeric]	Mean (sd) : 0.7 (1.2) min < med < max: 0 < 0 < 8 IQR (CV) : 1 (1.7)	7 distinct values	314 (100.0%)	0 (0.0%)
8	Parch [numeric]	Mean (sd) : 0.6 (1) min < med < max: 0 < 0 < 6 IQR (CV) : 1 (1.6)	7 distinct values	314 (100.0%)	0 (0.0%)
9	Ticket [character]	1. 347082 2. 2666 3. 110152 4. 113781 5. 13502 [242 others]	5 (1.6%) 4 (1.3%) 3 (1.0%) 3 (1.0%) 3 (1.0%) 296 (94.3%)	314 (100.0%)	0 (0.0%)
10	Fare [numeric]	Mean (sd) : 44.5 (58) min < med < max: 6.8 < 23 < 512.3 IQR (CV) : 42.9 (1.3)	156 distinct values	314 (100.0%)	0 (0.0%)
11	Cabin [character]	1. G6 2. E101 3. F33 4. B18 5. B28 [70 others]	4 (4.1%) 3 (3.1%) 3 (3.1%) 2 (2.1%) 2 (2.1%) 83 (85.6%)	97 (30.9%)	217 (69.1%)
12	Embarked [character]	1. C 2. Q 3. S	73 (23.4%) 36 (11.5%) 203 (65.1%)	312 (99.4%)	2 (0.6%)

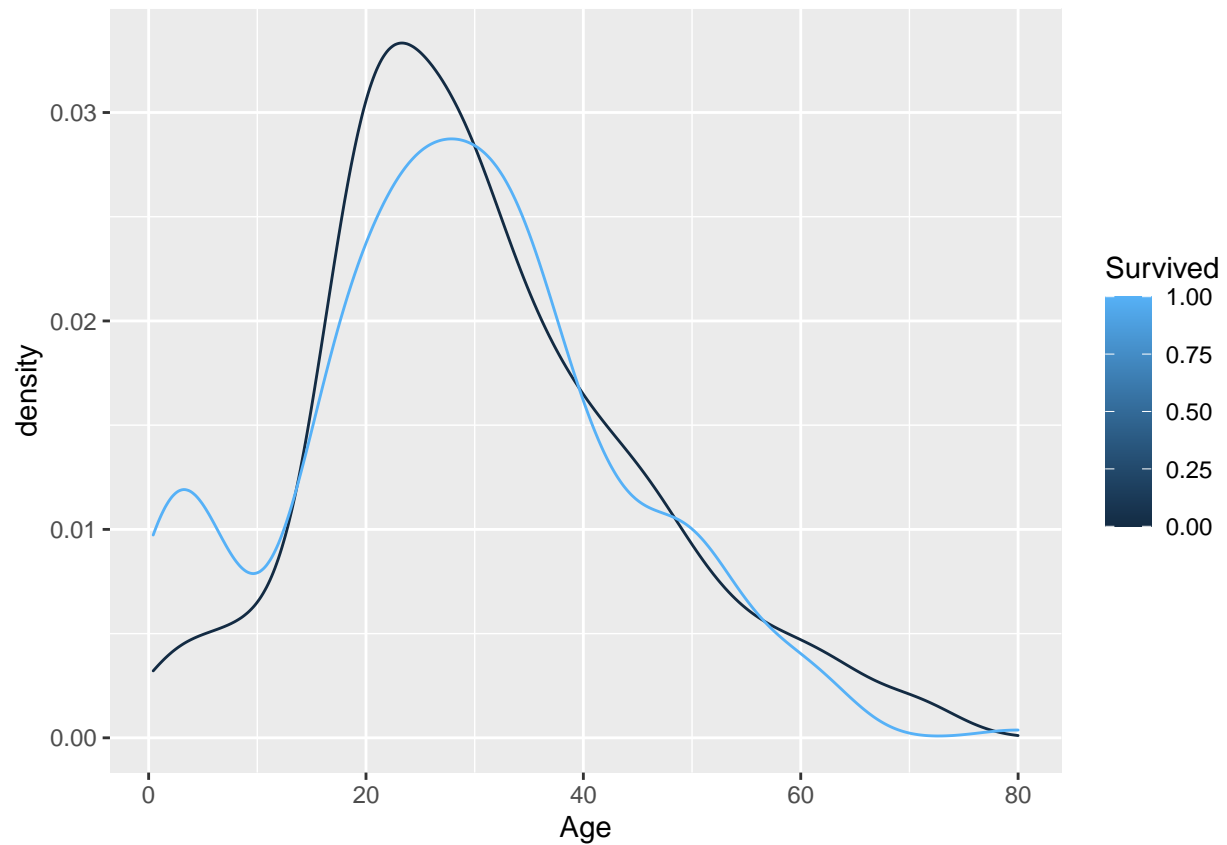**Group:** Sex = male
Dimensions: 577 x 12
Duplicates: 0

No	Variable	Stats / Values	Freqs (% of Valid)	Valid	Missing
1	PassengerId [numeric]	Mean (sd) : 454.1 (257.5) min < med < max: 1 < 464 < 891 IQR (CV) : 458 (0.6)	577 distinct values	577 (100.0%)	0 (0.0%)
2	Survived [numeric]	Min : 0 Mean : 0.2 Max : 1	0 : 468 (81.1%) 1 : 109 (18.9%)	577 (100.0%)	0 (0.0%)

No	Variable	Stats / Values	Freqs (% of Valid)	Valid	Missing
3	Pclass [numeric]	Mean (sd) : 2.4 (0.8) min < med < max: 1 < 3 < 3 IQR (CV) : 1 (0.3)	1 : 122 (21.1%) 2 : 108 (18.7%) 3 : 347 (60.1%)	577 (100.0%)	0 (0.0%)
4	Name [character]	1. Abbing, Mr. Anthony 2. Abbott, Mr. Rossmore Edwa 3. Abelson, Mr. Samuel 4. Adahl, Mr. Mauritz Nils M 5. Adams, Mr. John [572 others]	1 (0.2%) 1 (0.2%) 1 (0.2%) 1 (0.2%) 1 (0.2%) 572 (99.1%)	577 (100.0%)	0 (0.0%)
6	Age [numeric]	Mean (sd) : 30.7 (14.7) min < med < max: 0.4 < 29 < 80 IQR (CV) : 18 (0.5)	82 distinct values	453 (78.5%)	124 (21.5%)
7	SibSp [numeric]	Mean (sd) : 0.4 (1.1) min < med < max: 0 < 0 < 8 IQR (CV) : 0 (2.5)	7 distinct values	577 (100.0%)	0 (0.0%)
8	Parch [numeric]	Mean (sd) : 0.2 (0.6) min < med < max: 0 < 0 < 5 IQR (CV) : 0 (2.6)	6 distinct values	577 (100.0%)	0 (0.0%)
9	Ticket [character]	1. 1601 2. 3101295 3. S.O.C. 14879 4. 382652 5. CA 2144 [514 others]	7 (1.2%) 5 (0.9%) 5 (0.9%) 4 (0.7%) 4 (0.7%) 552 (95.7%)	577 (100.0%)	0 (0.0%)
10	Fare [numeric]	Mean (sd) : 25.5 (43.1) min < med < max: 0 < 10.5 < 512.3 IQR (CV) : 18.7 (1.7)	193 distinct values	577 (100.0%)	0 (0.0%)
11	Cabin [character]	1. F2 2. B51 B53 B55 3. B96 B98 4. C124 5. C23 C25 C27 [91 others]	3 (2.8%) 2 (1.9%) 2 (1.9%) 2 (1.9%) 2 (1.9%) 96 (89.7%)	107 (18.5%)	470 (81.5%)
12	Embarked [character]	1. C 2. Q 3. S	95 (16.5%) 41 (7.1%) 441 (76.4%)	577 (100.0%)	0 (0.0%)

1. ¿La varianza de las edades de quienes sobrevivieron es diferente para ambos grupos?
2. Con base en la respuesta anterior, prueba si las edades promedio son iguales o diferentes para quienes sobrevivieron o no?
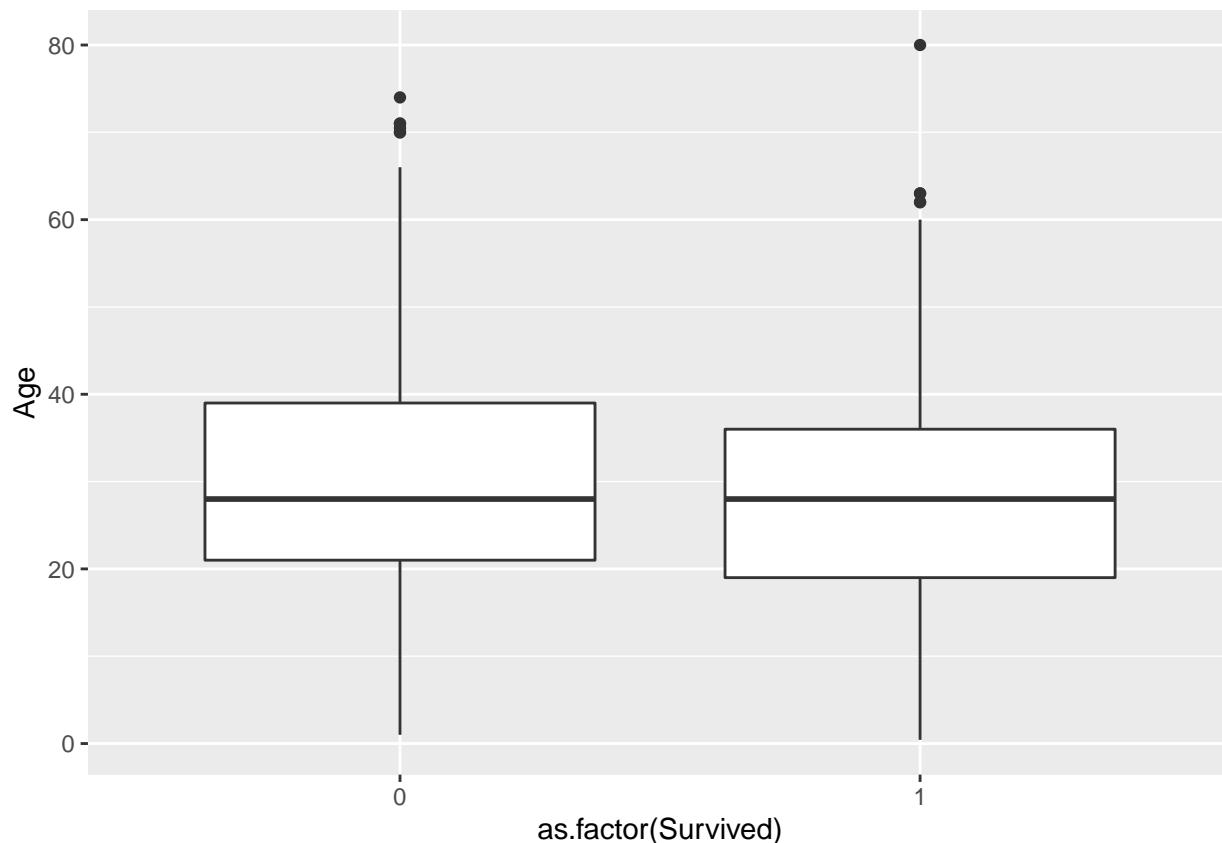
```
TitanicT %>% ggplot(aes(x=Age,colour=Survived,group=Survived)) +
  geom_density()
```

```
## Warning: Removed 177 rows containing non-finite values (stat_density).
```

```
TitanicT %>% ggplot(aes(x=as.factor(Survived), y=Age)) +
  geom_boxplot()
```

Warning: Removed 177 rows containing non-finite values (stat_boxplot).

```r
TitanicT %>% group_by(Survived) %>%  summarise(varAge = var(Age,na.rm = T))
```

```
## # A tibble: 2 x 2
##   Survived varAge
##      <dbl>  <dbl>
## 1        0   201.
## 2        1   224.
```

```r
var.test( Age ~ Survived, data = TitanicT, conf.level = 0.95)
```

```
##
##  F test to compare two variances
##
## data:  Age by Survived
## F = 0.89853, num df = 423, denom df = 289, p-value = 0.317
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.7253979 1.1082231
## sample estimates:
## ratio of variances
##           0.8985274
```

```r
t.test(Age ~ Survived, data = TitanicT, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  Age by Survived
## t = 2.046, df = 598.84, p-value = 0.04119
```

```
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  0.09158472 4.47339446
## sample estimates:
## mean in group 0 mean in group 1
##        30.62618        28.34369
```