STA 220 - Data and Web Technologies for Data Analysis - Lab 4

Consider the data set Titanic as follows:

`Description:`

This data set provides information on the fate of passengers on the fatal maiden voyage of the ocean liner 'Titanic', summarized according to economic status (class), sex, age and survival.

`Format:`

A 4-dimensional array resulting from cross-tabulating 2201 observations on 4 variables. The variables and their levels are as follows:

- Class with 4 levels: 1st, 2nd, 3rd, Crew.
- Sex with 2 levels: Male, Female.
- Age with 2 levels: Child, Adult.
- Survived with 2 levels: No, Yes.

For the data set Titanic Compute measures of frequency for the variables class and survived.

Hint

Use `apply(X, MARGIN, FUN, ...)`.

Solution

The data set

```
dim(Titanic)
## [1] 4 2 2 2
dimnames(Titanic)
## $Class
## [1] "1st"  "2nd"  "3rd"  "Crew"
##
## $Sex
## [1] "Male"   "Female"
##
## $Age
## [1] "Child" "Adult"
##
## $Survived
## [1] "No"  "Yes"
```

The variable Survived

```
sum(Titanic)
## [1] 2201
apply(Titanic, 4, sum)
##   No  Yes
## 1490  711
```

```
round(100 * apply(Titanic, "Survived", sum) / sum(Titanic), 1)
##   No  Yes
## 67.7 32.3
apply(Titanic, 1, sum)
##  1st  2nd  3rd Crew
##  325  285  706  885
```

The variables Survived and Class

```
apply(Titanic, c("Survived", "Class"), sum)
##          Class
## Survived 1st 2nd 3rd Crew
##      No  122 167 528  673
##      Yes 203 118 178  212
round(100 * apply(Titanic, c(4, 1), sum) / sum(Titanic), 1)
##          Class
## Survived 1st 2nd  3rd Crew
##      No  5.5 7.6 24.0 30.6
##      Yes 9.2 5.4  8.1  9.6
```