MG 221: Assignment II 2017

Due Date: **December 7, 2017**

The file `titanic.data`, which may be found in `http://www.mgmt.iisc.ernet.in/CM/MG221/Data_Files.html` contains observations on five variables pertaining to 1313 passengers on board the fateful Titanic voyage. The variables, which have been named in the first line of the file, are `Name`, `PClass`, `Age`, `Gender` and `Survived`. Among these five variables, apart from `Age`, the remaining four variables are qualitative or categorical in nature. We have no interest in the variable `Name`; the variable `PClass` refers to Passenger Class, which assumes three possible levels `1st`, `2nd` and `3rd`; the variable `Gender` has two levels `male` and `female`; and the 0-1 valued variable `Survived` assumes the value 1 if the passenger had survived and 0 otherwise. Based on this data set, answer the following:

1. Is there a significant difference in Age distribution between those who survived and those who did not?
2. Answer the same as above after controlling for Gender.
3. Remark on how Age affected the Survival Probability of a passenger on board the Titanic, based on consolidations of your findings in 1 and 2 above.
4. Is there a significant difference in Survival Probability between the two genders?
5. Is there a significant difference in Survival Probability between the three passenger classes?
6. Is there a significant difference in Survival Probability between the two genders even after taking the effect of Passenger Class into Account?

For answering all the questions above, first use an appropriate graphical technique, and then perform a formal test, after validating the necessary assumptions, wherever appropriate.