

MG 221: Assignment II 2020

Due Date: January 22, 2021

The file `titanic.data`, which may be found in http://www.mgmt.iisc.ac.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? [4]
2. Answer the same as above after controlling for Gender. [8]
3. Is there a significant difference in Survival Probabilities for the two Genders? [4]
4. Is there a significant difference in Survival Probabilities for the two Genders even after taking the effect of Passenger Class into account? [10]
5. Analyse how the Survival Probability got affected by the joint (combined) effect of Gender, Age and Passenger Class, and based on this analysis, write a brief and consolidated report on how the Survival Probabilities of Titanic passengers got affected by these three (independent) variables. [14]

For answering all the questions above, first use appropriate graphical techniques, and only then perform one or more formal statistical tests, after validating the necessary assumptions, wherever appropriate.