**Question 2:** Regress survival in next 6 months on comorbidities of the patients, age of patients, gender of patients and whether they participated in the medical foster home program. MFH is an intervention for nursing home patients.  In this program, nursing home patients are diverted to a community home and health care services are delivered within the community home.  The resident eats with the family and relies on the family members for socialization, food and comfort.  It is called "foster" home because the family previously living in the community home is supposed to act like the resident's family. Enrollment in MFH is indicated by a variable MFH=1.

Survival is reported in two variables.  One variable indicates survival in 6 months.  Another reports days known to survive, if the patient has died and otherwise null.  Thus a null value in this latter variable indicates the patient did not die.

CCS in these data refers to Clinical Classification System of Agency for Health Care Research and Quality.  These data indicate the comorbidities of the patient.  When null, it is assumed the patient did not have the comorbidity.  When data are entered it is assumed that the patient had the comorbidity and the reported value is the first (maximum) or last (minimum) number of days till admission to either the nursing home or the MFH. Thus an entry of 20 under the minimum CCS indicates that from the most recent occurrence of the comorbidity till admission was 20 days.  An entry of 400 under the Maximum CCS indicates that from the first time the comorbidity occurred till admission was 400 days. You choose what data (minimum, maximum, occurrence) is relevant for the analysis and you use what you think should be used. Keep in mind the possibility that for acute illness the most recent event may be predictive while for chronic illness the first occurrence may be predictive of cost.

The functional disabilities are probabilities that the patient has the disability.  These probabilities are generated from the CCS diagnoses and demographics of the person.

Clean the data using SQL. There are a number of cases that repeat and should be deleted from the analysis.  There are many null values.  The treatment of null value changes with the type of variable.  In some variables, null values indicate zero.  In others they can be estimated from the mode.  In still others, they should be treated as separate variable.  In completing this assignment follow these steps:

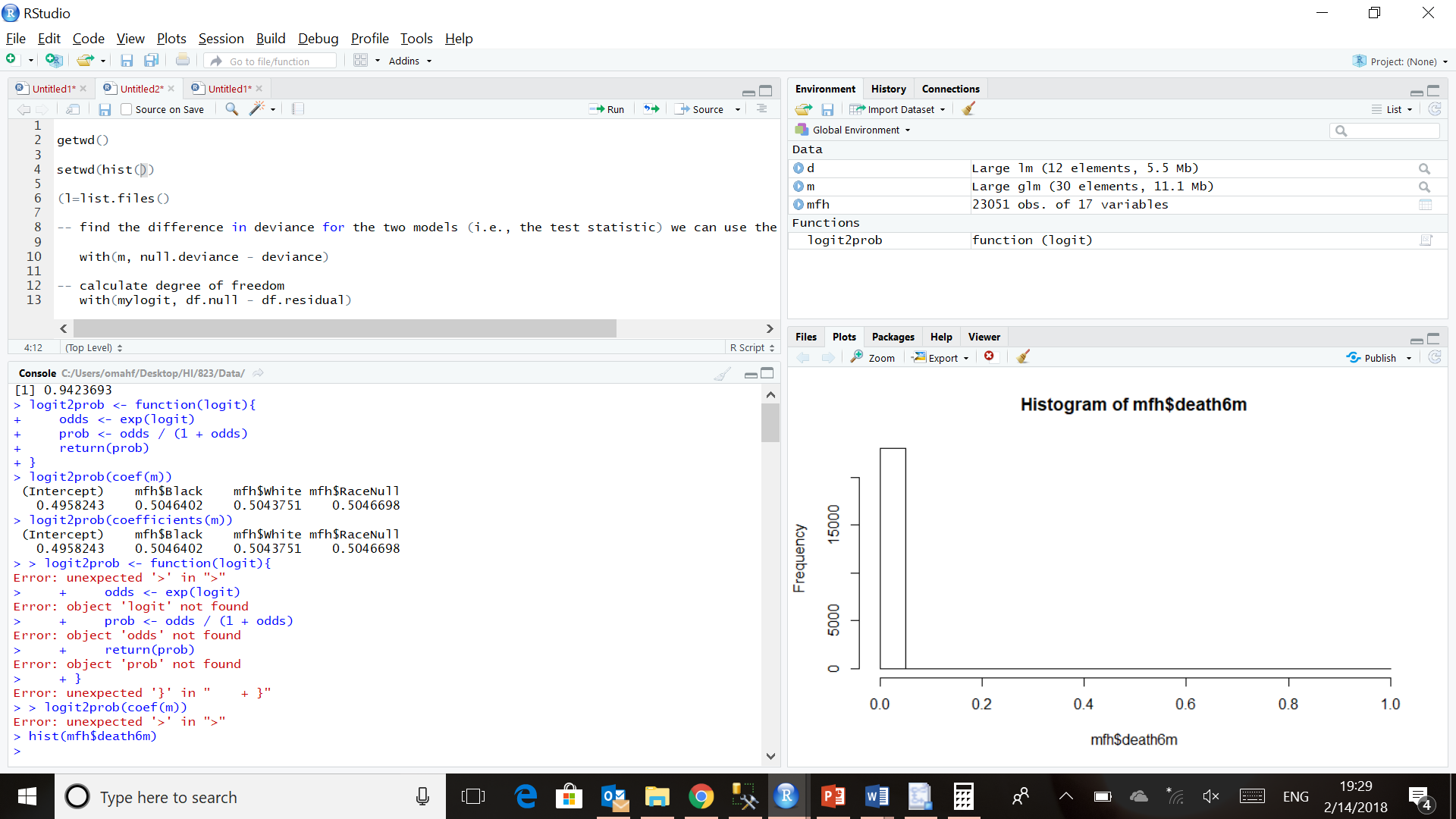
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variable | n | % | Mean | Std. Dev | Min | Max |
| Dead in 6 month |  |  |  |  |  |  |
| No(0) | 23,028 | 99.90 |  |  |  |  |
| Yes(1) | 23 | 0.10 |  |  |  |  |
| bathing |  |  |  |  |  |  |
| No(0) | 3,328 | 14.44 |  |  |  |  |
| Yes(1) | 19,724 | 85.56 |  |  |  |  |
| bladder |  |  |  |  |  |  |
| No(0) | 18,772 | 81.43 |  |  |  |  |
| Yes(1) | 4,280 | 18.57 |  |  |  |  |
| bowel |  |  |  |  |  |  |
| No(0) | 19,108 | 82.89 |  |  |  |  |
| Yes(1) | 3,944 | 17.11 |  |  |  |  |
| dressing |  |  |  |  |  |  |
| No(1) | 5,618 | 24.37 |  |  |  |  |
| Yes(1) | 17,434 | 75.63 |  |  |  |  |
| eating |  |  |  |  |  |  |
| No(0) | 14,300 | 62.03 |  |  |  |  |
| Yes(1) | 8,752 | 37.97 |  |  |  |  |
| Grooming |  |  |  |  |  |  |
| No(0) | 7,256 | 31.48 |  |  |  |  |
| Yes(1) | 15,796 | 68.52 |  |  |  |  |
| toileting |  |  |  |  |  |  |
| No(0) | 7,826 | 33.95 |  |  |  |  |
| Yes(1) | 15,226 | 66.05 |  |  |  |  |
| transferring |  |  |  |  |  |  |
| No(0) | 11,436 | 49.61 |  |  |  |  |
| Yes(1) | 11,616 | 49.61 |  |  |  |  |
| walking |  |  |  |  |  |  |
| No(0) | 3,413 | 14.81 |  |  |  |  |
| Yes(1) | 19,639 | 85.19 |  |  |  |  |
| MFH |  |  |  |  |  |  |
| No(0) | 22,216 | 96.37 |  |  |  |  |
| Yes(1) | 836 | 3.63 |  |  |  |  |
| Gender |  |  |  |  |  |  |
| Female(0) | 875 | 3.80 |  |  |  |  |
| Male(1) | 22,177 | 96.20 |  |  |  |  |
| Black |  |  |  |  |  |  |
| No(0) | 19,436 | 84.31 |  |  |  |  |
| Yes(1) | 3,616 | 15.69 |  |  |  |  |
| White |  |  |  |  |  |  |
| No(0) | 6,467 | 28.05 |  |  |  |  |
| Yes(1) | 16,585 | 71.95 |  |  |  |  |
| NO RAce |  |  |  |  |  |  |
| No(0) | 20,337 | 88.22 |  |  |  |  |
| Yes(1) | 2,715 | 11.78 |  |  |  |  |
| Age | 22,418 |  | 72.4 | 12.5 | 22 | 105 |

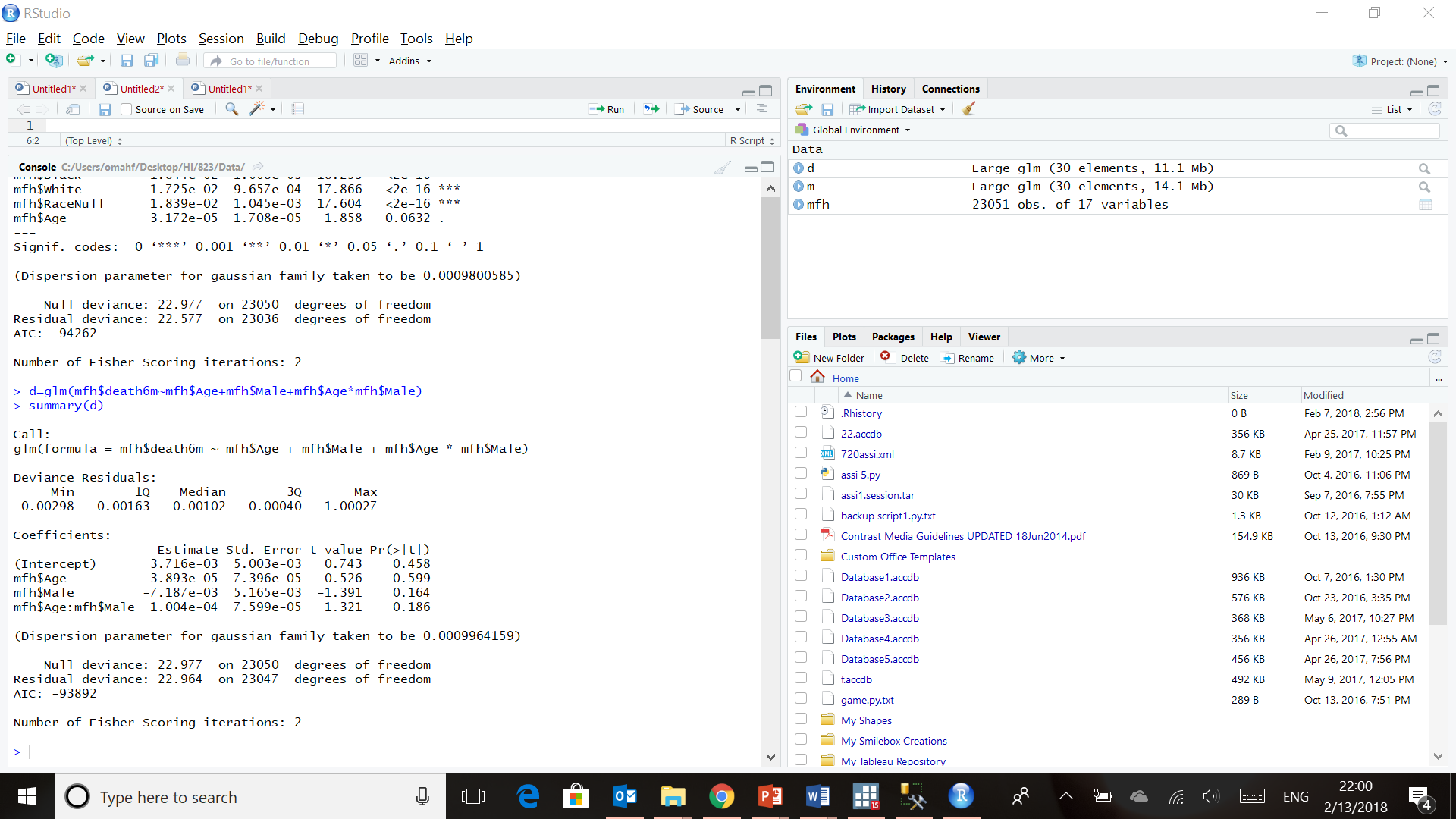
1. Describe the data using univariate analysis.

The table below shows characters distribution (n=23,051)

1. Check the distribution of the survival variable.

In this question our outcome variable is binary which has a binomial distribution.



1. Check the impact of the interaction of age and gender on survival.

From the result above we can see that there is no significant interaction between age and gender. In addition both variables show no significant impact on the outcome variable (survival).