**BS730**

**Project 1**

Diabetes is a major comorbidity for patients awaiting heart transplantation. In this study, we are interested in whether there are clinical or demographic differences at initial listing for heart transplant between patients with and without diabetes in a cohort of patients listed for heart transplant. Patients are listed on regional organ waiting lists and may wait up to 3 years or more before a heart becomes available. This study cohort also includes a subset of patients who received a transplant and we are interested in whether diabetes is associated with survival to hospital discharge following transplantation and any association between diabetes and change in clinical characteristics between listing and transplant.

**Project Guidelines**

You will create a Word document by using R MarkDown. Your report should address all of the specific questions below in Part II as well as include a brief Write-Up as outlined in Part III below. The write-up will include a statistical Methods section and a brief Results/Conclusion section summarizing the findings.

**Description of the dataset**

This dataset comes from the United Network for Organ Sharing (UNOS) which collects data on all transplant candidates and recipients in the United States. All patients listed for heart transplant between July 2006 and September 2010 are included. A subset of these patients received a heart transplant during this time period. Information was collected on clinical and demographic factors at the time of transplant as well as self-identified race/ethnicity categorized as White, Black, and Hispanic.

|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Description** | **Values** |
| *Demographic Factors and Insurance* | | |
| female | Female Sex | 1=Female  0=Male |
| medicaid | Medicaid Insurance | 0=Non-Medicaid Insurance 1=Medicaid Insurance |
| ethcat4 | Race/Ethnicity | 1=White 2=Black/African-American 3=Hispanic |
| *Diabetes and Cardiac Diagnosis* | | |
| diab4 | Diabetes | 1=No diabetes 2=Type I diabetes 3=Type II diabetes |
| diag | Diagnosis | 1=Cardiomyopathy 2=Ischemic 3=Congenital heart disease 4=Hypertrophic 5=Restrictive  6=Valvular 7=Other |
| *Clinical Characteristics at Listing* | | |
| age\_list | Age at Listing (years) | Continuous |
| bmi\_list | Body Mass Index at Listing | Continuous |
| gfr\_list | Glomular Filtration Rate at Listing, mL/min/1.73m2 | Continuous |
| *Clinical Characteristics at Transplant* | | |
| age\_tx | Age at Transplant (years) | Continuous |
| bmi\_tx | Body Mass Index at Transplant (kg/m2) | Continuous |
| gfr\_tx | Glomular Filtration Rate at Transplant, mL/min/1.73m2 | Continuous |
| pstatusdc | Survival Status at Discharge Following Heart Transplantation | 0=Alive 1=Dead |

**Part I.** Data manipulation (20 points)

* + 1. Read in the hearttransplant.csv data file in. Missing values are coded -999.
    2. Create a new variable called *agecat* that collapses *age\_list* into the following categories:

1=18-39 years

2=40-59 years

3=60-69 years

* + 1. Create a new variable called *diabetes* that is coded 1 for patients with diabetes (type I or type II) and coded 0 for patients without diabetes. Patients missing diabetes status should be coded missing on the new variable as well.
    2. Drop any patients with missing/unknown information on diabetes.
    3. Create a new variable called *tx\_complete* that is coded 1 if the patient actually received a heart transplant and 0 if they did not receive a transplant. Note that if the variable *pstatusdc* is missing, the patient did not receive a transplant.
    4. Create a new variable that is the change in GFR between listing (*gfr\_list*) and transplant (*gfr\_tx*) called *gfr\_delta*.
    5. Create a new variable that is the change in BMI between listing (*bmi\_list*) and transplant (*bmi\_tx*) called *bmi\_delta*.

**Part II.** Analyses (65 points)

For purposes of this project, you may assume that all continuous variables are normally distributed. Round all values to one or two decimal places, as needed.

1. (15 points) Are there differences in listing characteristics by diabetes status? Please use your newly created variable *diabetes* for the following analyses. You will create a descriptive table according to the template below with the appropriate summary statistics and p-values.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **No Diabetes**  **(N=xxx)** | **Diabetes**  **(N=xxx)** | **p-value** |
| **Age at Listing**, years |  |  |  |
| **Female Sex** |  |  |  |
| **Race/Ethnicity Group** |  |  |  |
| White |  |  |  |
| Black/African-American |  |  |
| Hispanic |  |  |
| **Medicaid Insurance** |  |  |  |
| **Diagnosis** |  |  |  |
| Cardiomyopathy |  |  |  |
| Ischemic |  |  |
| Congenital heart disease |  |  |
| Hypertrophic |  |  |
| Restrictive |  |  |
| Valvular |  |  |
| Other |  |  |
| **Body mass index at Listing, kg/m2** |  |  |  |
| **Glomular filtration rate at Listing, mL/min/1.73m2** |  |  |  |
| **Underwent Transplant** |  |  |  |

1. (10 points) Among the patients who received a transplant, is there an association between diabetes status and survival to discharge (*pstatusdc*)? Report the null and alternative hypothesis, test statistic, degrees of freedom, p-value, the appropriate measure of effect (with an interpretation), and conclusion.
2. (20 points). Now we will examine the change in BMI and GFR between listing and transplant separately among patients with and without diabetes. Derive the information required to fill in the following table. Note: the descriptive statistics should be calculated only among the subset of patients who have both measures for a given variable (e.g. if a patient is missing either bmi\_list or bmi\_tx, then they should be excluded from the analysis of change in BMI).

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **No Diabetes** | | | | | **Diabetes** | | | | |
| **N** | **Mean (SD)** | | | **p-value**  comparing listing  to transplant | **N** | **Mean (SD)** | | | **p-value**  comparing listing  to transplant |
| Listing | Transplant | Change | Listing | Transplant | Change |
| Body mass index, kg/m2 |  |  |  |  |  |  |  |  |  |  |
| Glomular Filtration Rate, mL/min/1.73m2 |  |  |  |  |  |  |  |  |  |  |

1. (10 points) Among patients with diabetes, was there a change in BMI or GFR between listing and transplant? Report the null and alternative hypothesis, test statistic, degrees of freedom, p-value and conclusion for each test.
2. (10 points) Among patients without diabetes, was there a change in BMI or GFR between listing and transplant? Report the null and alternative hypothesis, test statistic, degrees of freedom, p-value and conclusion for each test.

**Part III.** Write-Up (15 points)

* + - * 1. Write a brief statistical **Methods** section describing the analysis you performed in Part II. This will include the descriptive statistics used to summarize the data as well as the statistical methods (any hypothesis testing) used. For example: *We summarized continuous variables using median and 25%-75% percentiles, and categorical variables using counts and percentages. We used a two-sample t-test to assess the association between x and y*.
        2. Using the results from Part II, briefly summarize the findings with regard to the specific research questions being addressed in parts (a), (b) and (c) in a **Results/Conclusion** section. Comment upon interesting findings and summarize implications of your results. The entire section should be no more than one-half page long.