**BS723 Fall 2018**

**Project 2**

This analysis uses similar data to the dataset you used in Project 1 but we are interested only in patients who received a transplant. Additionally, you are given a secondary dataset containing more information on the heart donor and the volume of procedures (heart transplants) performed at the transplant center. We have two primary aims for this analysis. First, we would like to examine the association between change in glomerular filtration rate (GFR) between listing and transplant and the clinical/demographic characteristics at listing. Second, we would like to examine associations between survival to discharge after receiving a heart transplant and the clinical/demographic characteristics at listing, donor factors.

**Project Guidelines**

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 datasets**

For this project, you will need to use two datasets.

The first is provided as a csv dataset called **hearttransplant\_project2**and contains the same variables used in Project 1 along with the derived variables you were asked to create in Project 1 (*diabetes*, *agecat\_list*, *tx\_complete*, *gfr\_delta*, *bmi\_delta*). An additional variable called *pt\_code* is provided as a way of linking observations in this dataset to the observations in the second dataset described below.

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. This dataset contains the following variables:

|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Description** | **Values** |
| pt\_code | Unique Identifier | Numeric |
| *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 |
| whiterace | Race/Ethnicity | 0=Non-White (Black/African-American or Hispanic  1=White |
| *Diabetes and Cardiac Diagnosis* | | |
| diab4 | Diabetes | 1=No diabetes 2=Type I diabetes 3=Type II diabetes -9=Unknown/Missing |
| diabetes | Diabetes (created in Project 1) | 0=No Diabetes  1=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 |
| agecat\_list | Age Category at Listing (created in Project 1) | 1=18-39 years  2=40-59 years  3=60-79 years |
| bmi\_list | Body Mass Index at Listing | continuous |
| gfr\_list | Glomerular 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 | Glomerular Filtration Rate at Transplant ( mL/min/1.73m2) | continuous |
| pstatusdc | Survival Status at Discharge Following Heart Transplantation | 0=Alive 1=Dead |
| tx\_complete | Received transplant (created in Project 1) | 0=No  1=Yes |
|  |  |  |
| gfr\_delta | Change in Glomerular Filtration Rate between Listing and Transplant (mL/min/1.73m2; created in Project 1) | continuous |
| bmi\_delta | Change in BMI between Listing and Transplant (kg/m2; created in Project 1) | continuous |

The second is a CSV file called **transplant2.csv** and contains the following variables:

|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Description** | **Values** |
| pt\_code | Unique Identifier | Numeric |
| *Donor Characteristics* | | |
| age\_don | Donor Age (years) | continuous |
| gender\_don | Medicaid Insurance | F=Female  M=Male |
| *Transplant Center Volume* | | |
| listingvol | Number of heart transplant procedures performed at transplant center | continuous |

**Part I.** Data manipulation (25 points)

* 1. Create a data frame **transplant2** that is a copy of the dataset **hearttransplant\_project2** with the following changes:
     1. Restrict the dataset to only patients who received a heart transplant.
     2. Create a new diagnosis variable from *diag* called *diagnew* that is coded 1 for congenital heart disease, 2 for acquired conditions cardiomyopathy, ischemic, hypertrophic, restrictive and valvular diagnoses and 3 for other.
     3. Delete any observations with “other” diagnosis on the new variable *diagnew*.
  2. Import the CSV file **transplant2.csv** into a dataset called **donor**.
  3. Merge the newly imported dataset **donor** with the dataset **transplant2** and call the merged dataset **alltransplant**. NOTE: The two datasets should be merged by the unique patient identifier *pt\_code*.
  4. Using the merged dataset, we would like to create a new variable identifying a gender mismatch between the donor and recipient. If the donor and recipient are the same gender, the new variable *gender\_mismatch* should be coded 0 and if they are not the same the new variable should be coded 1.

**Part II.** Change in Glomerular Filtration Rate Analysis (30 points)

We would like to perform a regression analysis examining the association between change in GFR and the demographic and clinical characteristics of patients at listing.

1. Before constructing an adjusted regression model, we would like to perform a series of analyses examining whether diagnosis is a confounder of the association between diabetes and change in GFR.
   1. First we would like to see if there are differences in the change in glomerular filtration rate by the new diagnosis variable. Perform the appropriate test to examine the association between *gfr\_delta* and *diagnew*. Report the null and alternative hypotheses, test statistic, degrees of freedom, p-value and the appropriate measure of effect (with an interpretation) along with your conclusion. Comment on the equality of variances by diagnosis.
   2. Next perform the appropriate test examining the association between diagnosis (*diagnew*) and diabetes (*diabetes*). Report the null and alternative hypotheses, test statistic, degrees of freedom, p-value and conclusion.
   3. Based on the results in (1) and (2) above, is diagnosis potentially a confounder of the relationship between diabetes and change in GFR?
2. Next perform the appropriate regression analysis to predict change in GFR from diagnosis (*diagnew*), age at listing (*age\_list*), BMI at listing (*bmi\_list*), race (*whiterace*), gender (*female*) and diabetes (*diabetes*). Use the export\_summs function in jtools to produce a regression table similar to table 1 below. For continuous variables, derive the effect size and corresponding confidence interval associated with a one standard deviation change in the variable. Which variables are associated with change in GFR?

**Table 1**

|  |  |  |
| --- | --- | --- |
|  | **Parameter Estimate**  **(95% Confidence Interval)** | **p-value** |
| **Diagnosis** |  |  |
| Congenital |  |  |
| Acquired |  |  |
| **Age at Listing** |  |  |
| **BMI at Listing** |  |  |
| **Race** |  |  |
| White |  |  |
| Non-White |  |  |
| **Gender** |  |  |
| Male |  |  |
| Female |  |  |
| **Diabetes** |  |  |
| Yes |  |  |
| No |  |  |

**Part III.** Survival to Discharge Analysis (30 points)

1. Now we would like to examine the association between survival to discharge (*pstatusdc*) and the clinical and demographic characteristics at listing as well as the donor factors and transplant center volume. Perform the appropriate regression analysis with the following predictors: diagnosis (*diagnew*), age at listing (*age\_list*), BMI at listing (*bmi\_list*), white race (*whiterace*), gender (*female*), diabetes (*diabetes*), donor gender mismatch (*gender\_mismatch­*), donor age (*age\_don*) and center volume (*listingvol*). Use the export\_summs function in jtools to produce a regression table similar to table 2. For continuous variables, derive the effect size and corresponding confidence interval associated with a one standard deviation change in the variable. Which predictors are associated with survival to hospital discharge following a heart transplant?
2. Perform a test to examine whether donor gender mismatch is an effect modifier of the association between diabetes and survival to discharge in a model adjusting for the same variables as (a). Fully report your test of interaction. Can you conclude that gender mismatch is an effect modifier of the relationship between diabetes and survival to discharge?

**Table 2**

|  |  |  |
| --- | --- | --- |
|  | **Odds Ratio**  **(95% Confidence Interval)** | **p-value** |
| **Diagnosis** |  |  |
| Congenital |  |  |
| Acquired |  |  |
| **Age at Listing** |  |  |
| **BMI at Listing** |  |  |
| **Race** |  |  |
| White |  |  |
| Non-White |  |  |
| **Gender** |  |  |
| Male |  |  |
| Female |  |  |
| **Diabetes** |  |  |
| Yes |  |  |
| No |  |  |
| **Gender Mismatch** |  |  |
| Yes |  |  |
| No |  |  |
| **Donor Age** |  |  |
| **Center Volume** |  |  |

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

* + - * 1. Write a brief (no more than one-half page) statistical **Methods** section describing the analysis you performed in Parts II and III. This will include the descriptive statistics used to summarize the data as well as the statistical methods (any hypothesis testing or regression analysis) used. For example: *We used a two-sample t-test when assessing the association between x and y*. Do not give specific procedures nor a full list of all analyses you did for each variable.
        2. Using the results from Parts II and III, briefly summarize the findings with regard to the specific research questions being addressed 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.