Project 2

Garima Lohani

6 December 2018

### Part 1a

ds<-read.table("hearttransplant\_project2.csv", header=TRUE, stringsAsFactors=FALSE,sep=",")  
str(ds)

## 'data.frame': 9929 obs. of 19 variables:  
## $ pt\_code : int 725820 790529 663806 691112 717512 779159 783446 691227 740970 687009 ...  
## $ age\_tx : int 58 NA NA 57 55 60 69 NA 53 56 ...  
## $ age\_list : int 58 62 57 57 54 60 68 52 52 56 ...  
## $ pstatusdc : int 0 NA NA 0 0 0 0 NA 0 0 ...  
## $ female : int 0 0 1 1 1 1 0 0 0 0 ...  
## $ ethcat4 : int 2 1 2 1 1 2 1 1 1 1 ...  
## $ gfr\_list : num 50.4 25.1 47 86.2 39.2 ...  
## $ gfr\_tx : num 41.8 NA NA 86.2 46.6 ...  
## $ medicaid : int 0 0 1 0 0 0 0 0 0 0 ...  
## $ diag : int 2 1 2 1 1 1 6 1 1 6 ...  
## $ diab4 : int 1 1 3 1 1 3 1 1 1 3 ...  
## $ bmi\_list : num 23.6 30.3 26.8 29.5 32.6 ...  
## $ bmi\_tx : num 22.7 30.8 26.8 29.5 32.1 ...  
## $ agecat\_list: int 2 3 2 2 2 3 3 2 2 2 ...  
## $ diabetes : int 0 0 1 0 0 1 0 0 0 1 ...  
## $ tx\_complete: int 1 0 0 1 1 1 1 0 1 1 ...  
## $ gfr\_delta : num -8.62 NA NA 0 7.46 ...  
## $ bmi\_delta : num -9.12e-01 4.70e-01 0.00 -3.78e-05 -5.08e-01 ...  
## $ whiterace : int 0 1 0 1 1 0 1 1 1 1 ...

transplant2<-ds  
transplant2<-subset(ds,!ds$tx\_complete==0)  
table(transplant2$tx\_complete)

##   
## 1   
## 6362

table(transplant2$diag)

##   
## 1 2 3 4 5 6 7   
## 3098 2243 150 138 152 146 435

transplant2$diagnew<-3  
transplant2$diagnew[transplant2$diag==3]<-1  
transplant2$diagnew[transplant2$diag==1 | transplant2$diag==2 |transplant2$diag==4|transplant2$diag==5 | transplant2$diag==6 ]<-2  
  
table(transplant2$diagnew)

##   
## 1 2 3   
## 150 5777 435

transplant2<-subset(transplant2,!transplant2$diagnew==3)  
  
table(transplant2$diagnew)

##   
## 1 2   
## 150 5777

### Part 1b

donor<-read.table("transplant2.csv", header=TRUE, stringsAsFactors=FALSE,sep=",")

### Part 1c

alltransplant<-merge(transplant2,donor,by="pt\_code")  
  
str(alltransplant)

## 'data.frame': 5927 obs. of 23 variables:  
## $ pt\_code : int 3426 4920 7936 10919 13999 18737 23596 32987 39963 40294 ...  
## $ age\_tx : int 61 58 66 62 51 59 69 66 58 41 ...  
## $ age\_list : int 61 58 65 61 51 59 69 66 58 41 ...  
## $ pstatusdc : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ female : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ ethcat4 : int 3 1 1 1 3 1 1 2 1 1 ...  
## $ gfr\_list : num 56.1 62.2 55.4 85.8 78.8 ...  
## $ gfr\_tx : num 38.6 76.7 46.8 88.9 89 ...  
## $ medicaid : int 0 0 0 0 0 0 0 0 0 1 ...  
## $ diag : int 1 1 2 2 1 2 1 1 2 1 ...  
## $ diab4 : int 1 1 3 3 1 2 1 1 2 1 ...  
## $ bmi\_list : num 22.2 34.4 24.6 31.6 28.9 ...  
## $ bmi\_tx : num 21.7 32.9 25.9 30.6 29 ...  
## $ agecat\_list: int 3 2 3 3 2 2 3 3 2 2 ...  
## $ diabetes : int 0 0 1 1 0 1 0 0 1 0 ...  
## $ tx\_complete: int 1 1 1 1 1 1 1 1 1 1 ...  
## $ gfr\_delta : num -17.57 14.56 -8.58 3.13 10.18 ...  
## $ bmi\_delta : num -0.5149 -1.505 1.2941 -1.0108 0.0592 ...  
## $ whiterace : int 0 1 1 1 0 1 1 0 1 1 ...  
## $ diagnew : num 2 2 2 2 2 2 2 2 2 2 ...  
## $ age\_don : int 19 59 41 42 19 23 25 47 47 36 ...  
## $ gender\_don : chr "M" "M" "M" "M" ...  
## $ listingvol : int 58 194 75 142 29 71 149 67 120 43 ...

str(alltransplant$female)

## int [1:5927] 0 0 0 0 0 0 0 0 0 0 ...

### Part 1d

table(alltransplant$female, alltransplant$gender\_don)

##   
## F M  
## 0 846 3600  
## 1 781 700

alltransplant$gender\_mismatch<- 1  
alltransplant$gender\_mismatch[(alltransplant$female==1 & alltransplant$gender\_don=='F')| (alltransplant$female==0 & alltransplant$gender\_don=='M')]<- 0  
   
str(alltransplant$gender\_mismatch)

## num [1:5927] 0 0 0 0 0 0 0 0 0 0 ...

### Part 2ai

ttest2 <- t.test(alltransplant$gfr\_delta ~ alltransplant$diagnew)  
ttest2

##   
## Welch Two Sample t-test  
##   
## data: alltransplant$gfr\_delta by alltransplant$diagnew  
## t = 1.1491, df = 145.82, p-value = 0.2524  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -5.338416 20.169465  
## sample estimates:  
## mean in group 1 mean in group 2   
## 8.545283 1.129758

ftest2 <- var.test(alltransplant$gfr\_delta[alltransplant$diagnew == 1], alltransplant$gfr\_delta[alltransplant$diagnew == 2])  
ftest2

##   
## F test to compare two variances  
##   
## data: alltransplant$gfr\_delta[alltransplant$diagnew == 1] and alltransplant$gfr\_delta[alltransplant$diagnew == 2]  
## F = 9.1059, num df = 145, denom df = 5695, p-value < 2.2e-16  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 7.30217 11.65721  
## sample estimates:  
## ratio of variances   
## 9.105855

### I think two sample t test is appropriate for this case.  
###Null Hypotheses H0: The mean of change in glomerular filtration rate in cogenital diagnosis is same as the mean of change in glomerular filtration rate in acquired diagnosis.  
###Alternative Hypotheses HA: The mean of change in glomerular filtration rate in cogenital diagnosis is different from the mean of change in glomerular filtration rate in acquired diagnosis.  
### The test statistics is 1.1491, degree of freedom is 145.82 and pvalue is 0.2524.  
### Since the p value is less than alpha value of 0.05, we reject null hypotheses and conclude that there is significant relationship between diagnew and gfr\_delta  
  
###The variance are not equal.They are far off. As the p value of variance test is less than the alpha value of 0.05.

### Part 2aii

library(epitools)

## Warning: package 'epitools' was built under R version 3.4.2

library(Hmisc)

## Warning: package 'Hmisc' was built under R version 3.4.4

## Loading required package: lattice

## Loading required package: survival

##   
## Attaching package: 'survival'

## The following object is masked from 'package:epitools':  
##   
## ratetable

## Loading required package: Formula

## Warning: package 'Formula' was built under R version 3.4.4

## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 3.4.4

##   
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':  
##   
## format.pval, units

c1<-table(alltransplant$diagnew,alltransplant$diabetes)  
colnames(c1)<-c("No diabetes","Diabetes")  
rownames(c1)<-c("Cogenital", "Acquired")  
  
#performing chi-square test of independence  
res<-chisq.test(c1,correct=FALSE)  
res$expected

##   
## No diabetes Diabetes  
## Cogenital 112.7468 37.25325  
## Acquired 4342.2532 1434.74675

res

##   
## Pearson's Chi-squared test  
##   
## data: c1  
## X-squared = 29.248, df = 1, p-value = 6.369e-08

oddsratio(t(c1),method="wald")

## $data  
##   
## Cogenital Acquired Total  
## No diabetes 141 4314 4455  
## Diabetes 9 1463 1472  
## Total 150 5777 5927  
##   
## $measure  
## odds ratio with 95% C.I.  
## estimate lower upper  
## No diabetes 1.000000 NA NA  
## Diabetes 5.313012 2.701225 10.45011  
##   
## $p.value  
## two-sided  
## midp.exact fisher.exact chi.square  
## No diabetes NA NA NA  
## Diabetes 9.882157e-10 1.265997e-09 6.369029e-08  
##   
## $correction  
## [1] FALSE  
##   
## attr(,"method")  
## [1] "Unconditional MLE & normal approximation (Wald) CI"

### The chi squared test is appropriate for this case as we are comparing are two categorical variable.  
  
###Null Hypotheses H0: the true probabilities of event are independent in the diagnew and diabetes.  
  
###ALternative Hypotheses H1: the true probabilities of event are dependent in the diabetes and diagnew.  
  
### The test statistics is 29.48, the degree of freedom is 1 and p value is 6.369e-08  
### Since p-value is significant at 0.05 alpha value, we reject the null hypotheses and there exists evidence that diagnew and diabetes are associated to each other.

### Part 2aiii

###Yes it is possible that diagnosis is a confounder of the relationship between diabetesand gfr\_delta

### Part 2b

### 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).

linreg <- lm(gfr\_delta~diagnew+age\_list+bmi\_list+whiterace+female+diabetes,   
 data = alltransplant)  
summary(linreg)

##   
## Call:  
## lm(formula = gfr\_delta ~ diagnew + age\_list + bmi\_list + whiterace +   
## female + diabetes, data = alltransplant)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -262.38 -10.93 -1.02 10.48 855.22   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 21.52056 5.15912 4.171 3.07e-05 \*\*\*  
## diagnew -5.09557 2.42690 -2.100 0.03580 \*   
## age\_list -0.10347 0.03201 -3.232 0.00123 \*\*   
## bmi\_list -0.20670 0.07502 -2.755 0.00588 \*\*   
## whiterace 1.36320 0.83420 1.634 0.10228   
## female -0.05292 0.86916 -0.061 0.95145   
## diabetes -0.72241 0.89325 -0.809 0.41870   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 28.24 on 5834 degrees of freedom  
## (86 observations deleted due to missingness)  
## Multiple R-squared: 0.005637, Adjusted R-squared: 0.004614   
## F-statistic: 5.512 on 6 and 5834 DF, p-value: 1.053e-05

### Use the export\_summs function in jtools to produce a regression table similar to table 1 below

library(jtools)

## Warning: package 'jtools' was built under R version 3.4.4

export\_summs(linreg,error\_format='[{conf.low},{conf.high},p={p.value}]',scale=TRUE,stars.pvalue=FALSE)

## Warning in knit\_print.huxtable(x, ...): Falling back to markdown for Word  
## document. Many features will not be supported.

## Warning in to\_md.huxtable(ht = structure(list(names = c("",  
## "(Intercept)", : Markdown cannot handle cells with colspan/rowspan > 1

## Warning in to\_md.huxtable(ht = structure(list(names = c("",  
## "(Intercept)", : Can't vary column alignment in markdown; using first row

|  |  |
| --- | --- |
|  | Model 1 |
| (Intercept) | 5.51 \* |
|  | [0.52,10.50,p=0.03] |
| diagnew | -5.10 \* |
|  | [-9.85,-0.34,p=0.04] |
| age\_list | -1.28 \*\* |
|  | [-2.05,-0.50,p=0.00] |
| bmi\_list | -1.05 \*\* |
|  | [-1.79,-0.30,p=0.01] |
| whiterace | 1.36 |
|  | [-0.27,3.00,p=0.10] |
| female | -0.05 |
|  | [-1.76,1.65,p=0.95] |
| diabetes | -0.72 |
|  | [-2.47,1.03,p=0.42] |
| N | 5841 |
| R2 | 0.01 |
| \*\*\* p < 0.001; \*\* p < | 0.01; \* p < 0.05. |
| |  |  |  | | --- | --- | --- | |  | **Parameter Estimate**  **(95% Confidence Interval)** | **p-value** | | **Diagnosis** | -5.10 | 0.04 | | Congenital |  |  | | Acquired |  |  | | **Age at Listing** | -1.28 | 0.00 | | **BMI at Listing** | -1.05 | 0.01 | | **Race** | 1.36 | 0.10 | | White |  |  | | Non-White |  |  | | **Gender** | -0.05 | 0.95 | | Male |  |  | | Female |  |  | | **Diabetes** | -0.72 | 0.42 | | Yes |  |  | | No |  |  | |  |

### For continuous variables, derive the effect size and corresponding confidence interval associated with a one standard deviation change in the variable.

###age\_list= [-9.85,-0.34,p=0.04] and -1.28  
### bmi\_list = [-2.05,-0.50,p=0.00] and -1.05

### Which variables are associated with change in GFR?

summary(linreg)

##   
## Call:  
## lm(formula = gfr\_delta ~ diagnew + age\_list + bmi\_list + whiterace +   
## female + diabetes, data = alltransplant)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -262.38 -10.93 -1.02 10.48 855.22   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 21.52056 5.15912 4.171 3.07e-05 \*\*\*  
## diagnew -5.09557 2.42690 -2.100 0.03580 \*   
## age\_list -0.10347 0.03201 -3.232 0.00123 \*\*   
## bmi\_list -0.20670 0.07502 -2.755 0.00588 \*\*   
## whiterace 1.36320 0.83420 1.634 0.10228   
## female -0.05292 0.86916 -0.061 0.95145   
## diabetes -0.72241 0.89325 -0.809 0.41870   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 28.24 on 5834 degrees of freedom  
## (86 observations deleted due to missingness)  
## Multiple R-squared: 0.005637, Adjusted R-squared: 0.004614   
## F-statistic: 5.512 on 6 and 5834 DF, p-value: 1.053e-05

###diagnew, age\_list and bmi\_list are the only variables which look to be associated with gfr\_delta.

### Part 3a

### 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).

logreg = glm(pstatusdc ~ diagnew+age\_list+bmi\_list+whiterace+female+  
 diabetes+gender\_mismatch+age\_don+listingvol,   
 data = alltransplant, family = binomial)  
summary(logreg)

##   
## Call:  
## glm(formula = pstatusdc ~ diagnew + age\_list + bmi\_list + whiterace +   
## female + diabetes + gender\_mismatch + age\_don + listingvol,   
## family = binomial, data = alltransplant)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.0053 -0.3536 -0.3036 -0.2618 2.8339   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.7025190 0.6008749 -2.833 0.00461 \*\*   
## diagnew -1.7220959 0.2570143 -6.700 2.08e-11 \*\*\*  
## age\_list 0.0263203 0.0058056 4.534 5.80e-06 \*\*\*  
## bmi\_list 0.0092243 0.0116134 0.794 0.42704   
## whiterace -0.1879470 0.1328484 -1.415 0.15714   
## female 0.0314871 0.1392492 0.226 0.82111   
## diabetes -0.1443382 0.1430213 -1.009 0.31287   
## gender\_mismatch 0.2590429 0.1312703 1.973 0.04846 \*   
## age\_don 0.0213089 0.0048048 4.435 9.21e-06 \*\*\*  
## listingvol -0.0004901 0.0006326 -0.775 0.43844   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2438.6 on 5925 degrees of freedom  
## Residual deviance: 2364.0 on 5916 degrees of freedom  
## (1 observation deleted due to missingness)  
## AIC: 2384  
##   
## Number of Fisher Scoring iterations: 6

### Use the export\_summs function in jtools to produce a regression table similar to table 2.

export\_summs(logreg,error\_format='[{conf.low},{conf.high}, ,p={p.value}]',scale=TRUE,stars.pvalue=FALSE)

## Warning in knit\_print.huxtable(x, ...): Falling back to markdown for Word  
## document. Many features will not be supported.

## Warning in to\_md.huxtable(ht = structure(list(names = c("",  
## "(Intercept)", : Markdown cannot handle cells with colspan/rowspan > 1

## Warning in to\_md.huxtable(ht = structure(list(names = c("",  
## "(Intercept)", : Can't vary column alignment in markdown; using first row

|  |  |
| --- | --- |
|  | Model 1 |
| (Intercept) | -1.23 \*\*\* |
|  | [-1.78,-0.68, ,p=0.00] |
| diagnew | -1.72 \*\*\* |
|  | [-2.23,-1.22, ,p=0.00] |
| age\_list | 0.32 \*\*\* |
|  | [0.18,0.46, ,p=0.00] |
| bmi\_list | 0.05 |
|  | [-0.07,0.16, ,p=0.43] |
| whiterace | -0.19 |
|  | [-0.45,0.07, ,p=0.16] |
| female | 0.03 |
|  | [-0.24,0.30, ,p=0.82] |
| diabetes | -0.14 |
|  | [-0.42,0.14, ,p=0.31] |
| gender\_mismatch | 0.26 \* |
|  | [0.00,0.52, ,p=0.05] |
| age\_don | 0.25 \*\*\* |
|  | [0.14,0.36, ,p=0.00] |
| listingvol | -0.05 |
|  | [-0.16,0.07, ,p=0.44] |
| N | 5926 |
| AIC | 2384.03 |
| BIC | 2450.90 |
| Pseudo R2 | 0.04 |
| \*\*\* p < 0.001; \*\* p | < 0.01; \* p < 0.05. |
| |  |  |  | | --- | --- | --- | |  | **Odds Ratio**  **(95% Confidence Interval)** | **p-value** | | **Diagnosis** | -1.72 | 0.00 | | Congenital |  |  | | Acquired |  |  | | **Age at Listing** | 0.32 | 0.00 | | **BMI at Listing** | 0.05 | 0.43 | | **Race** | -0.19 | 0.16 | | White |  |  | | Non-White |  |  | | **Gender** | 0.03 | 0.82 | | Male |  |  | | Female |  |  | | **Diabetes** | -0.14 | 0.31 | | Yes |  |  | | No |  |  | | **Gender Mismatch** | 0.26 | 0.05 | | Yes |  |  | | No |  |  | | **Donor Age** | 0.25 | 0.00 | | **Center Volume** | -0.05 | 0.44 | |  |

### For continuous variables, derive the effect size and corresponding confidence interval associated with a one standard deviation change in the variable

###age\_list 0.32 and [0.18,0.46]  
###bmi\_list 0.05 and [-0.07,0.16]   
###age\_don 0.25 and [0.14,0.36]   
###listingvol -0.05 and [-0.16,0.07]

### Which predictors are associated with survival to hospital discharge following a heart transplant?

summary(logreg)

##   
## Call:  
## glm(formula = pstatusdc ~ diagnew + age\_list + bmi\_list + whiterace +   
## female + diabetes + gender\_mismatch + age\_don + listingvol,   
## family = binomial, data = alltransplant)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.0053 -0.3536 -0.3036 -0.2618 2.8339   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.7025190 0.6008749 -2.833 0.00461 \*\*   
## diagnew -1.7220959 0.2570143 -6.700 2.08e-11 \*\*\*  
## age\_list 0.0263203 0.0058056 4.534 5.80e-06 \*\*\*  
## bmi\_list 0.0092243 0.0116134 0.794 0.42704   
## whiterace -0.1879470 0.1328484 -1.415 0.15714   
## female 0.0314871 0.1392492 0.226 0.82111   
## diabetes -0.1443382 0.1430213 -1.009 0.31287   
## gender\_mismatch 0.2590429 0.1312703 1.973 0.04846 \*   
## age\_don 0.0213089 0.0048048 4.435 9.21e-06 \*\*\*  
## listingvol -0.0004901 0.0006326 -0.775 0.43844   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2438.6 on 5925 degrees of freedom  
## Residual deviance: 2364.0 on 5916 degrees of freedom  
## (1 observation deleted due to missingness)  
## AIC: 2384  
##   
## Number of Fisher Scoring iterations: 6

###diagnew, age\_list, age\_don and gender\_mismatch are the variables that are significant at 5% aplha level

### Part 3b

### 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

logreg2 = glm(pstatusdc ~ diagnew + age\_list + bmi\_list + whiterace +   
 female + diabetes + age\_don + listingvol + diabetes\*gender\_mismatch,   
 data = alltransplant, family = binomial)  
summary(logreg2)

##   
## Call:  
## glm(formula = pstatusdc ~ diagnew + age\_list + bmi\_list + whiterace +   
## female + diabetes + age\_don + listingvol + diabetes \* gender\_mismatch,   
## family = binomial, data = alltransplant)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.0073 -0.3536 -0.3037 -0.2618 2.8348   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.7041463 0.6010851 -2.835 0.00458 \*\*   
## diagnew -1.7230601 0.2572660 -6.698 2.12e-11 \*\*\*  
## age\_list 0.0263360 0.0058091 4.534 5.80e-06 \*\*\*  
## bmi\_list 0.0092464 0.0116100 0.796 0.42579   
## whiterace -0.1881275 0.1328641 -1.416 0.15679   
## female 0.0316821 0.1392726 0.227 0.82005   
## diabetes -0.1366836 0.1683219 -0.812 0.41677   
## age\_don 0.0213119 0.0048047 4.436 9.18e-06 \*\*\*  
## listingvol -0.0004891 0.0006327 -0.773 0.43944   
## gender\_mismatch 0.2650092 0.1484220 1.786 0.07418 .   
## diabetes:gender\_mismatch -0.0259198 0.3018192 -0.086 0.93156   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2438.6 on 5925 degrees of freedom  
## Residual deviance: 2364.0 on 5915 degrees of freedom  
## (1 observation deleted due to missingness)  
## AIC: 2386  
##   
## Number of Fisher Scoring iterations: 6

###After adding the interaction term between gender\_mismatch and diabetes, the signs of coefficients do not change. Hence gender\_mismatch is not an effect modifier

#### Part 4a

### We used two-sample t-test to examine the association between gfr\_delta and diagnew.  
  
  
###We performed Chi- square test of independence to examine the association between diagnosis (diagnew) and diabetes (diabetes).   
  
###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), we performed simple linear regression.  
  
### The logistic regression is used 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.   
  
###We performed logistic regression 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 used previously.   
  
### P-values are considered significant at the 0.05 level.

###From the results of Part 2a, there is significant relationship between diagnosis and glomerular filtration rate with p value of 0.2524 .Also,there exists evidence that diagnosis and diabetes are associated to each other with significant p value of 6.369e-08.We need to further investigate if diagnosis is a confounder of the relationship between diabetes and glomerular filtration rate.  
  
### From the results of Part 2b, diagnosis, age at lising and bmi at listing are the only variables which are associated with glomerular filtration rate with significantly low p values.  
  
### From the results of Part 3a, diagnosis, age atlisting, age of donor and gender mismatch of receipient and donor are the variables that are significant at 5% aplha level and are associated with survival to hospital discharge following a heart transplant  
  
### From the results of Part 3b, we conclude that gender\_mismatch is not an effect modifier. After adding the interaction term between gender\_mismatch and diabetes, the signs of coefficients of do not change.