Project 5

Anthony Jayce Forgeron (B00738840)

06/12/2020

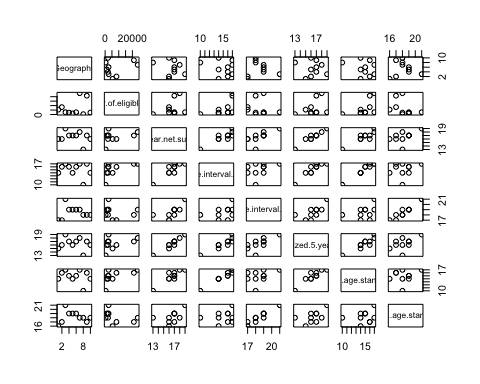
Here is where we will collect the packages and load the libraries that we will use for plotting throughout the project (this part has been hidden to save space but does still run)

Here we are going to be reading in all of the data that has been collected

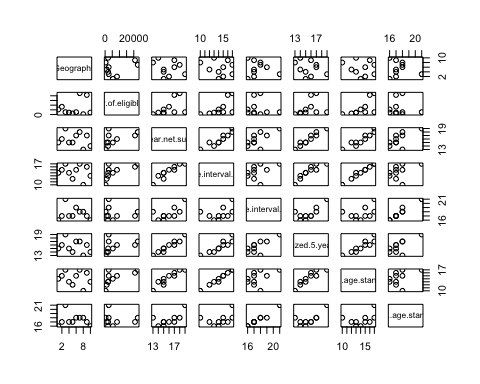
#reading in data from the year 2008 to 2010  
data1 = read.csv("project1.csv")  
#reading in data from the year 2007 to 2009  
data2 = read.csv("project2.csv")  
#reading in data from the year 2006 to 2008  
data3 = read.csv("project3.csv")

Now we will move on to creating plots comparing all data categories to each other so that we can get a better look at any potential relations in linearity

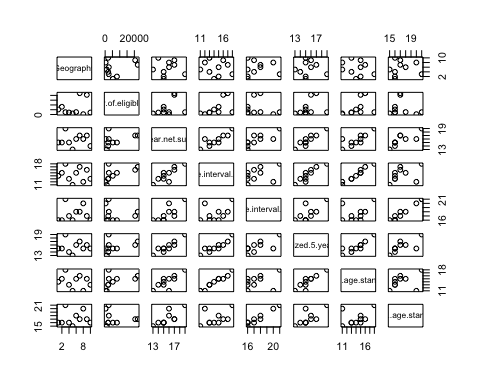
plot(data1)



plot(data2)

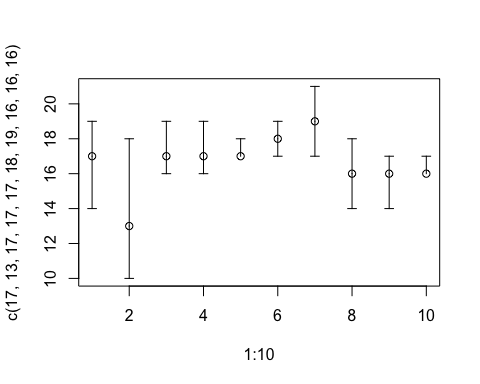


plot(data3)

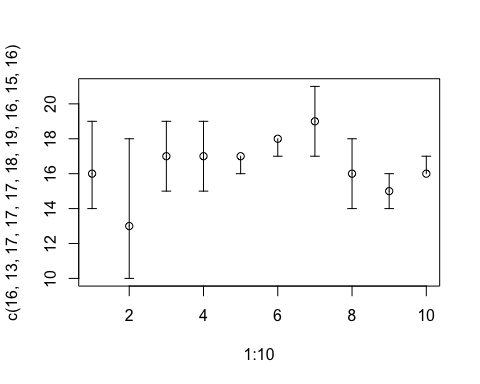
 The common theme that can be observe from above is that there is a almost completely linear relation that is taking place between the X5 year net survival and age standardize year net survival (this is probably becuase just in general they are closely related in type and collection)

Moving onto confidence interval observation we are going to be making multiple data tables plotting the various data with their confidence intervals shown in error bars

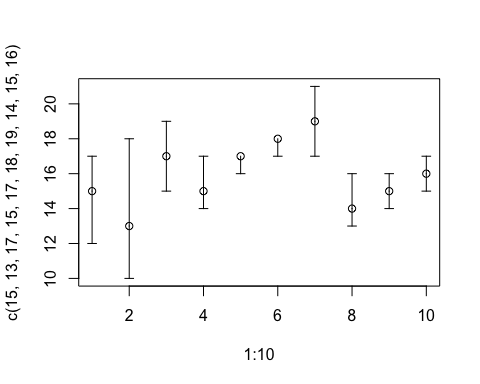
#CI from 2008 to 2010 for X5  
plotCI(data1$X5.year.net.survival, ui = data1$High.95..confidence.interval..5.year.net.survival, li = data1$Low.95..confidence.interval..5.year.net.survival)



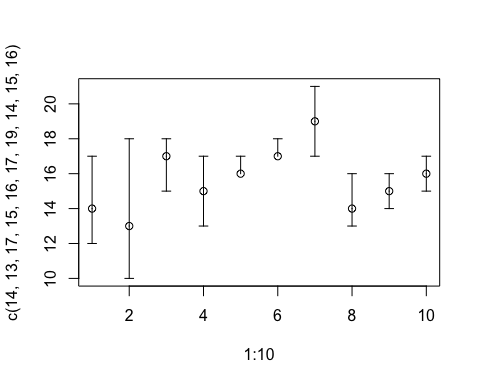
#CI from 2008 to 2010 for age standardized  
plotCI(data1$Age.standardized.5.year.net.survival, ui = data1$High.95..confidence.interval..age.standardized.5.year.net.survival, li = data1$Low.95..confidence.interval..age.standardized.5.year.net.survival)



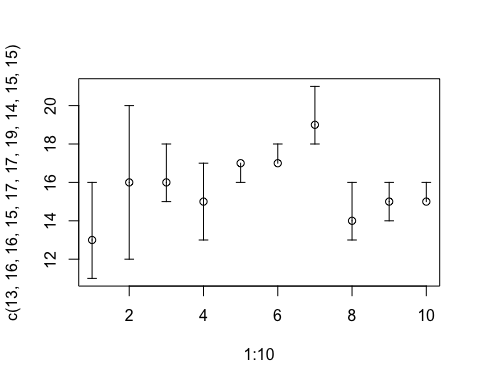
#CI from 2007 to 2009 for X5  
plotCI(data2$X5.year.net.survival, ui = data2$High.95..confidence.interval..5.year.net.survival, li = data2$Low.95..confidence.interval..5.year.net.survival)



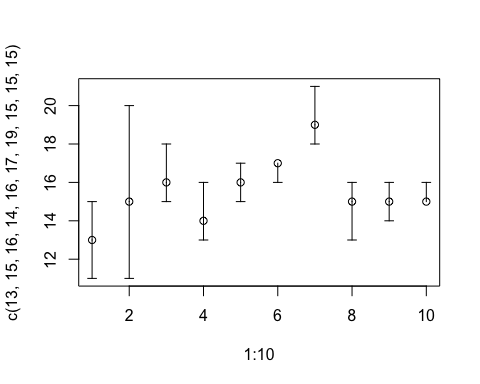
#CI from 2007 to 2009 for age standardized  
plotCI(data2$Age.standardized.5.year.net.survival, ui = data2$High.95..confidence.interval..age.standardized.5.year.net.survival, li = data2$Low.95..confidence.interval..age.standardized.5.year.net.survival)



#CI for 2006 to 2008 for X5  
plotCI(data3$X5.year.net.survival, ui = data3$High.95..confidence.interval..5.year.net.survival, li = data3$Low.95..confidence.interval..5.year.net.survival)



#CI for 2006 to 2008 for age standardized  
plotCI(data3$Age.standardized.5.year.net.survival, ui = data3$High.95..confidence.interval..age.standardized.5.year.net.survival, li = data3$Low.95..confidence.interval..age.standardized.5.year.net.survival)

 One obvious trend that can be observed would be that when we move from the X5 category over to the age standardized category we can see the confidence intervals becoming widers (this is valid for all datasets)

What we are going to take a look at now is some linear models that we should be taking into consideration based off the plots that were made in the beginning

lm1 = lm(data1$Number.of.eligible.cases~data1$X5.year.net.survival+data1$Age.standardized.5.year.net.survival+data1$X5.year.net.survival\*data1$Age.standardized.5.year.net.survival)  
lm2 = lm(data2$Number.of.eligible.cases~data2$X5.year.net.survival+data2$Age.standardized.5.year.net.survival+data2$X5.year.net.survival\*data2$Age.standardized.5.year.net.survival)  
lm3 = lm(data3$Number.of.eligible.cases~data3$X5.year.net.survival+data3$Age.standardized.5.year.net.survival+data3$X5.year.net.survival\*data3$Age.standardized.5.year.net.survival)  
first = summary(lm1)  
first

##   
## Call:  
## lm(formula = data1$Number.of.eligible.cases ~ data1$X5.year.net.survival +   
## data1$Age.standardized.5.year.net.survival + data1$X5.year.net.survival \*   
## data1$Age.standardized.5.year.net.survival)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7267 -6606 -1375 2413 14136   
##   
## Coefficients:  
## Estimate  
## (Intercept) -121455.7  
## data1$X5.year.net.survival 2521.2  
## data1$Age.standardized.5.year.net.survival 12035.8  
## data1$X5.year.net.survival:data1$Age.standardized.5.year.net.survival -403.7  
## Std. Error  
## (Intercept) 236322.9  
## data1$X5.year.net.survival 14201.9  
## data1$Age.standardized.5.year.net.survival 19460.2  
## data1$X5.year.net.survival:data1$Age.standardized.5.year.net.survival 937.9  
## t value  
## (Intercept) -0.514  
## data1$X5.year.net.survival 0.178  
## data1$Age.standardized.5.year.net.survival 0.618  
## data1$X5.year.net.survival:data1$Age.standardized.5.year.net.survival -0.430  
## Pr(>|t|)  
## (Intercept) 0.626  
## data1$X5.year.net.survival 0.865  
## data1$Age.standardized.5.year.net.survival 0.559  
## data1$X5.year.net.survival:data1$Age.standardized.5.year.net.survival 0.682  
##   
## Residual standard error: 9468 on 6 degrees of freedom  
## Multiple R-squared: 0.177, Adjusted R-squared: -0.2346   
## F-statistic: 0.43 on 3 and 6 DF, p-value: 0.7391

second = summary(lm2)  
second

##   
## Call:  
## lm(formula = data2$Number.of.eligible.cases ~ data2$X5.year.net.survival +   
## data2$Age.standardized.5.year.net.survival + data2$X5.year.net.survival \*   
## data2$Age.standardized.5.year.net.survival)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -10858 -2712 1436 3572 6172   
##   
## Coefficients:  
## Estimate  
## (Intercept) -160417.8  
## data2$X5.year.net.survival 19083.7  
## data2$Age.standardized.5.year.net.survival 188.3  
## data2$X5.year.net.survival:data2$Age.standardized.5.year.net.survival -554.6  
## Std. Error  
## (Intercept) 164126.1  
## data2$X5.year.net.survival 9865.7  
## data2$Age.standardized.5.year.net.survival 12615.3  
## data2$X5.year.net.survival:data2$Age.standardized.5.year.net.survival 644.6  
## t value  
## (Intercept) -0.977  
## data2$X5.year.net.survival 1.934  
## data2$Age.standardized.5.year.net.survival 0.015  
## data2$X5.year.net.survival:data2$Age.standardized.5.year.net.survival -0.860  
## Pr(>|t|)  
## (Intercept) 0.366  
## data2$X5.year.net.survival 0.101  
## data2$Age.standardized.5.year.net.survival 0.989  
## data2$X5.year.net.survival:data2$Age.standardized.5.year.net.survival 0.423  
##   
## Residual standard error: 6261 on 6 degrees of freedom  
## Multiple R-squared: 0.6151, Adjusted R-squared: 0.4226   
## F-statistic: 3.195 on 3 and 6 DF, p-value: 0.1051

third = summary(lm3)  
third

##   
## Call:  
## lm(formula = data3$Number.of.eligible.cases ~ data3$X5.year.net.survival +   
## data3$Age.standardized.5.year.net.survival + data3$X5.year.net.survival \*   
## data3$Age.standardized.5.year.net.survival)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8687 -4559 -1926 4039 11832   
##   
## Coefficients:  
## Estimate  
## (Intercept) -238770.0  
## data3$X5.year.net.survival 15088.1  
## data3$Age.standardized.5.year.net.survival 14067.0  
## data3$X5.year.net.survival:data3$Age.standardized.5.year.net.survival -852.6  
## Std. Error  
## (Intercept) 214343.0  
## data3$X5.year.net.survival 12976.8  
## data3$Age.standardized.5.year.net.survival 15268.0  
## data3$X5.year.net.survival:data3$Age.standardized.5.year.net.survival 831.2  
## t value  
## (Intercept) -1.114  
## data3$X5.year.net.survival 1.163  
## data3$Age.standardized.5.year.net.survival 0.921  
## data3$X5.year.net.survival:data3$Age.standardized.5.year.net.survival -1.026  
## Pr(>|t|)  
## (Intercept) 0.308  
## data3$X5.year.net.survival 0.289  
## data3$Age.standardized.5.year.net.survival 0.392  
## data3$X5.year.net.survival:data3$Age.standardized.5.year.net.survival 0.345  
##   
## Residual standard error: 8430 on 6 degrees of freedom  
## Multiple R-squared: 0.2768, Adjusted R-squared: -0.08484   
## F-statistic: 0.7654 on 3 and 6 DF, p-value: 0.5536

the models cannot be adjusted anymore than how they already are since there is no implication that we need to perform a backwarde elimination with an critical value of 0.05

Now we will perform an overall AnovA across all of the different years for the data

one = anova(lm1)  
one

## Analysis of Variance Table  
##   
## Response: data1$Number.of.eligible.cases  
## Df  
## data1$X5.year.net.survival 1  
## data1$Age.standardized.5.year.net.survival 1  
## data1$X5.year.net.survival:data1$Age.standardized.5.year.net.survival 1  
## Residuals 6  
## Sum Sq  
## data1$X5.year.net.survival 69427078  
## data1$Age.standardized.5.year.net.survival 29614441  
## data1$X5.year.net.survival:data1$Age.standardized.5.year.net.survival 16610146  
## Residuals 537902617  
## Mean Sq  
## data1$X5.year.net.survival 69427078  
## data1$Age.standardized.5.year.net.survival 29614441  
## data1$X5.year.net.survival:data1$Age.standardized.5.year.net.survival 16610146  
## Residuals 89650436  
## F value  
## data1$X5.year.net.survival 0.7744  
## data1$Age.standardized.5.year.net.survival 0.3303  
## data1$X5.year.net.survival:data1$Age.standardized.5.year.net.survival 0.1853  
## Residuals   
## Pr(>F)  
## data1$X5.year.net.survival 0.4127  
## data1$Age.standardized.5.year.net.survival 0.5864  
## data1$X5.year.net.survival:data1$Age.standardized.5.year.net.survival 0.6819  
## Residuals

two = anova(lm2)  
two

## Analysis of Variance Table  
##   
## Response: data2$Number.of.eligible.cases  
## Df  
## data2$X5.year.net.survival 1  
## data2$Age.standardized.5.year.net.survival 1  
## data2$X5.year.net.survival:data2$Age.standardized.5.year.net.survival 1  
## Residuals 6  
## Sum Sq  
## data2$X5.year.net.survival 155885427  
## data2$Age.standardized.5.year.net.survival 190889629  
## data2$X5.year.net.survival:data2$Age.standardized.5.year.net.survival 29013951  
## Residuals 235199112  
## Mean Sq  
## data2$X5.year.net.survival 155885427  
## data2$Age.standardized.5.year.net.survival 190889629  
## data2$X5.year.net.survival:data2$Age.standardized.5.year.net.survival 29013951  
## Residuals 39199852  
## F value  
## data2$X5.year.net.survival 3.9767  
## data2$Age.standardized.5.year.net.survival 4.8697  
## data2$X5.year.net.survival:data2$Age.standardized.5.year.net.survival 0.7402  
## Residuals   
## Pr(>F)   
## data2$X5.year.net.survival 0.09318 .  
## data2$Age.standardized.5.year.net.survival 0.06946 .  
## data2$X5.year.net.survival:data2$Age.standardized.5.year.net.survival 0.42265   
## Residuals   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

three = anova(lm3)  
three

## Analysis of Variance Table  
##   
## Response: data3$Number.of.eligible.cases  
## Df  
## data3$X5.year.net.survival 1  
## data3$Age.standardized.5.year.net.survival 1  
## data3$X5.year.net.survival:data3$Age.standardized.5.year.net.survival 1  
## Residuals 6  
## Sum Sq  
## data3$X5.year.net.survival 85909228  
## data3$Age.standardized.5.year.net.survival 2486274  
## data3$X5.year.net.survival:data3$Age.standardized.5.year.net.survival 74763838  
## Residuals 426339841  
## Mean Sq  
## data3$X5.year.net.survival 85909228  
## data3$Age.standardized.5.year.net.survival 2486274  
## data3$X5.year.net.survival:data3$Age.standardized.5.year.net.survival 74763838  
## Residuals 71056640  
## F value  
## data3$X5.year.net.survival 1.2090  
## data3$Age.standardized.5.year.net.survival 0.0350  
## data3$X5.year.net.survival:data3$Age.standardized.5.year.net.survival 1.0522  
## Residuals   
## Pr(>F)  
## data3$X5.year.net.survival 0.3137  
## data3$Age.standardized.5.year.net.survival 0.8578  
## data3$X5.year.net.survival:data3$Age.standardized.5.year.net.survival 0.3446  
## Residuals