

Math 150 - Methods in Biostatistics - Project 1

Barah Makhdum

Due: Monday, April 8, 2019, in class

```
library(tidyverse)
library(broom)
library(dplyr)
```

```
AIDSdata <- read_csv("AIDSdata.csv")
dim(AIDSdata)
```

```
## [1] 851 16
```

we have Sample size N is equal 851 and 16 variables

The Time by day and the range between one to 362 days. The median of time is 257 days and the mean is equal 231.8

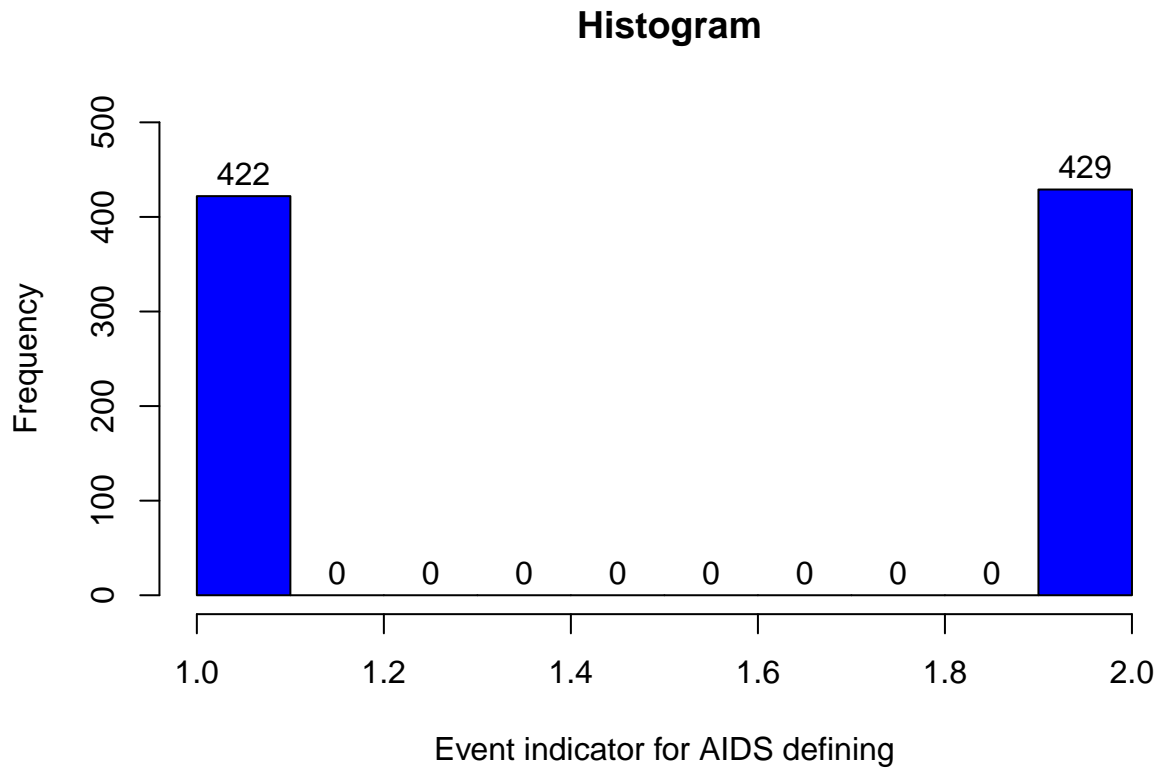
```
summary(AIDSdata$time)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       1.0   179.5   257.0   231.8   300.0   362.0
```

The variable censor is a nominal variable (categorical variable). There are 782 people who said (1) for AIDS defining diagnosis or death and 69 who choose (0) for Otherwise

The tx is a categorical variable define as a two groups. First for 422 people who take the treatment includes IDV and the second 429 for who follow treatment regimen without IDV.

```
hist(AIDSdata$txgrp, col="blue", xlab="Event indicator for AIDS defining", main=" Histogram",ylim =c(0,
```



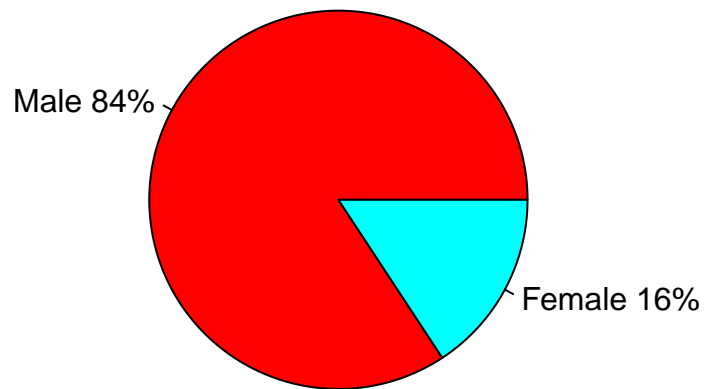
From the Histogram we can describe the variable txgrp we have 4 treatment groups. However, there are only two treatment groups used in this study. which are (ZDV + 3TC) and (ZDV + 3TC + IDV).

The strat2 is a CD4 stratum at screening and it was divided into two groups. zero was for whom had CD4 less than or equal to 50 and one given to whom had greater than 50 at CD4 screening

```
matrix(c("CD4<=50",sum(AIDSdata$strat2==0),"CD4>50",sum(AIDSdata$strat2==1)),ncol=2,byrow=TRUE)
```

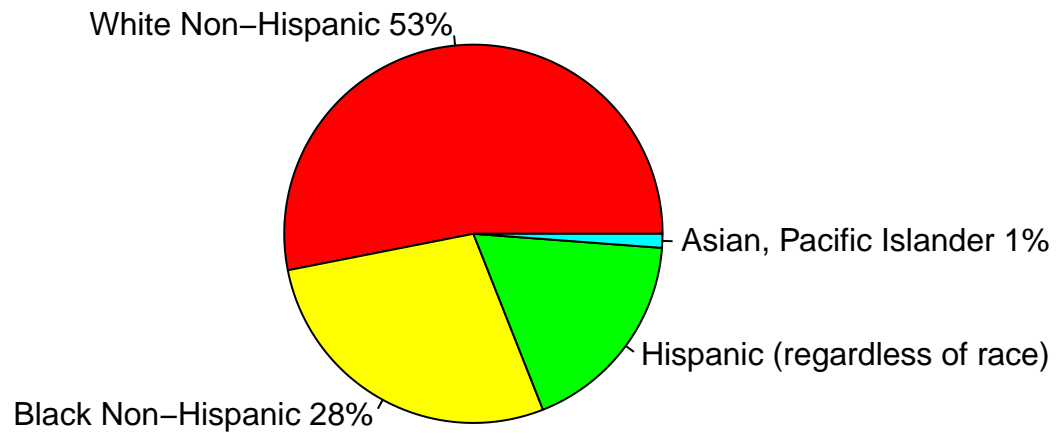
```
##      [,1]      [,2]
## [1,] "CD4<=50" "327"
## [2,] "CD4>50"  "524"
```

```
a<-sum(AIDSdata$sex==1)
b<-sum(AIDSdata$sex==2)
slices<-c(a,b)
labs<-c("Male","Female")
pct<-round(slices/sum(slices)*100)
lbls<-paste(labs,pct)
lbls<-paste(lbls,"%",sep="")
pie(slices,lbls,col=rainbow(length(lbls)))
```



From the graph above, we have 84% Male in the study and 16% Female.

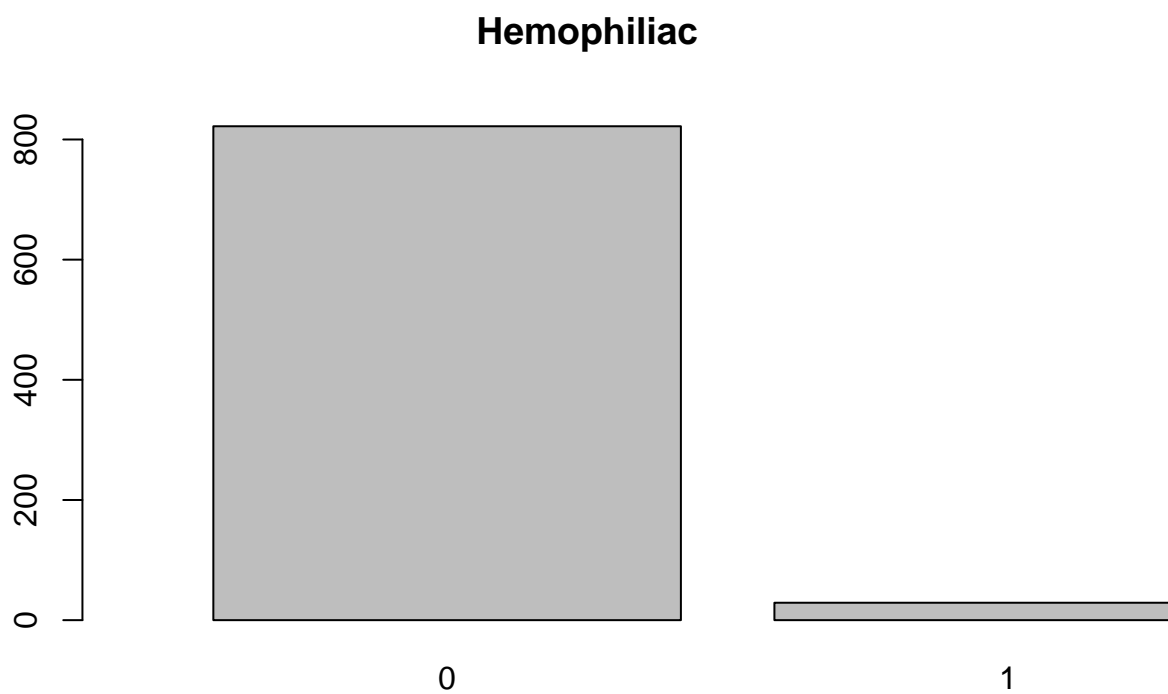
```
a<-sum(AIDSdata$raceth==1)
b<-sum(AIDSdata$raceth==2)
c<-sum(AIDSdata$raceth==3)
d<-sum(AIDSdata$raceth==4)
slices<-c(a,b,c,d)
labs<-c("White Non-Hispanic","Black Non-Hispanic","Hispanic (regardless of race)","Asian, Pacific Islander")
pct<-round(slices/sum(slices)*100)
lbls<-paste(labs,pct)
lbls<-paste(lbls,"%",sep="")
pie(slices,lbls,col=rainbow(length(lbls)))
```



This graph show the results of the study in which Race they are. White (Non-Hispanic) is highest percentage. Hispanic and Black percentage are close to each other. However, Asian, took the lowest percentage.

ivdrug was the question about the IV drug use history and the choices were (never, currently and previously).

```
counts <- table(AIDSdata$hemophil)  
barplot(counts, xlim=c(0,2), ylim = c(0,850),main="Hemophilic")
```



Summarise the information in the graph by showing the number of said No they do not have Hemophiliac greater than to whom said yes they have it.