

## Exercise 2-2

Mehrab Atighi

4/15/2021

### Introducing data to R

```
rm(list = ls())  
#now we want to attach our data to R from notepad or txt.  
Heart <- read.table("F://lessons//regression2//exercise//2-2//New Text  
Document.txt", sep=",", head=T, row.names=1)  
#now we want to see some of our data  
head(Heart)  
  
##   sbp tobacco  ldl adiposity famhist typea obesity alcohol age chd  
## 1 160   12.00 5.73   23.11 Present   49   25.30   97.20 52   1  
## 2 144    0.01 4.41   28.61 Absent    55   28.87    2.06 63   1  
## 3 118    0.08 3.48   32.28 Present   52   29.14    3.81 46   0  
## 4 170    7.50 6.41   38.03 Present   51   31.99   24.26 58   1  
## 5 134   13.60 3.50   27.78 Present   60   25.99   57.34 49   1  
## 6 132    6.20 6.47   36.21 Present   62   30.77   14.14 45   0  
  
#the fifth column of our dataset or famhist is not numeric and we get Labe (  
present = 1 ,Absent = 0)  
Heart[,5]=ifelse(Heart[,5]=="Absent",0,1)
```

sbp column indicates systolic blood pressure.

tobacco column indicates cumulative tobacco (Kg).

ldl column indicates low density lipoprotein cholesterol.

adiposity column it is a concept of body masses that are often in the form of fat.

famhist column indicates family history of heart disease (Present, Absent).

typea column indicates type-A behavior.

obesity column it is a concept of obesity that has different types and according to its coefficient, different information can be obtained.

alcohol column indicates current alcohol consumption.

age column indicates age at onset.

chd column indicates response, coronary heart disease.

## Solve

chek correlation between the variables:

cor(Heart)

```
##          sbp      tobacco      ldl      adiposity      famhist
## sbp      1.00000000  0.21224652  0.15829633  0.35650008  0.08564531
## tobacco  0.21224652  1.00000000  0.15890546  0.28664037  0.08860143
## ldl      0.15829633  0.15890546  1.00000000  0.44043175  0.16135306
## adiposity 0.35650008  0.28664037  0.44043175  1.00000000  0.18172101
## famhist  0.08564531  0.08860143  0.16135306  0.18172101  1.00000000
## typea    -0.05745431 -0.01460788  0.04404758 -0.04314364  0.04480858
## obesity   0.23806661  0.12452941  0.33050586  0.71655625  0.11559508
## alcohol   0.14009559  0.20081339 -0.03340340  0.10033013  0.08051969
## age       0.38877060  0.45033016  0.31179923  0.62595442  0.23966742
## chd       0.19235411  0.29971754  0.26305268  0.25412139  0.27237273
##          typea      obesity      alcohol      age      chd
## sbp      -0.05745431  0.23806661  0.14009559  0.3887706  0.19235411
## tobacco  -0.01460788  0.12452941  0.20081339  0.4503302  0.29971754
## ldl      0.04404758  0.33050586 -0.03340340  0.3117992  0.26305268
## adiposity -0.04314364  0.71655625  0.10033013  0.6259544  0.25412139
## famhist  0.04480858  0.11559508  0.08051969  0.2396674  0.27237273
## typea    1.00000000  0.07400610  0.03949794 -0.1026063  0.10315583
## obesity   0.07400610  1.00000000  0.05161957  0.2917771  0.10009508
## alcohol   0.03949794  0.05161957  1.00000000  0.1011246  0.06253068
## age      -0.10260632  0.29177713  0.10112465  1.0000000  0.37297334
## chd       0.10315583  0.10009508  0.06253068  0.3729733  1.00000000
```

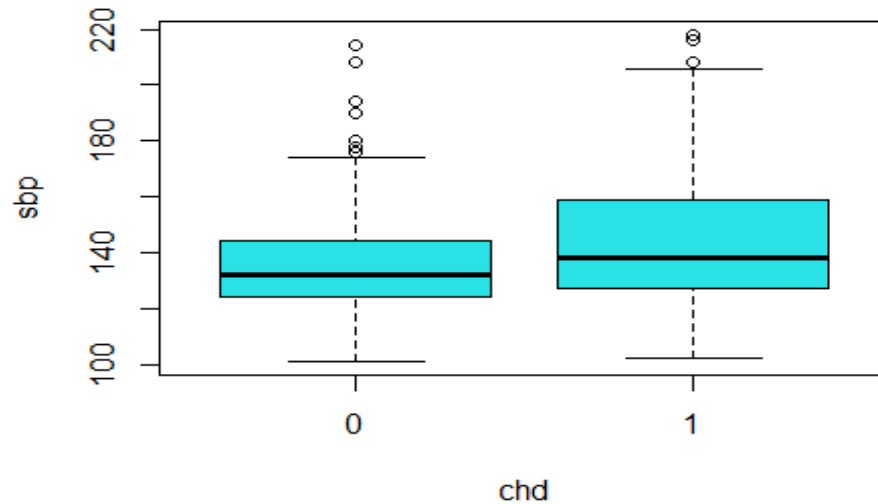
check the relation between response and variables with Boxplots:

*#at the first we should attach the data*

```
attach(Heart)
```

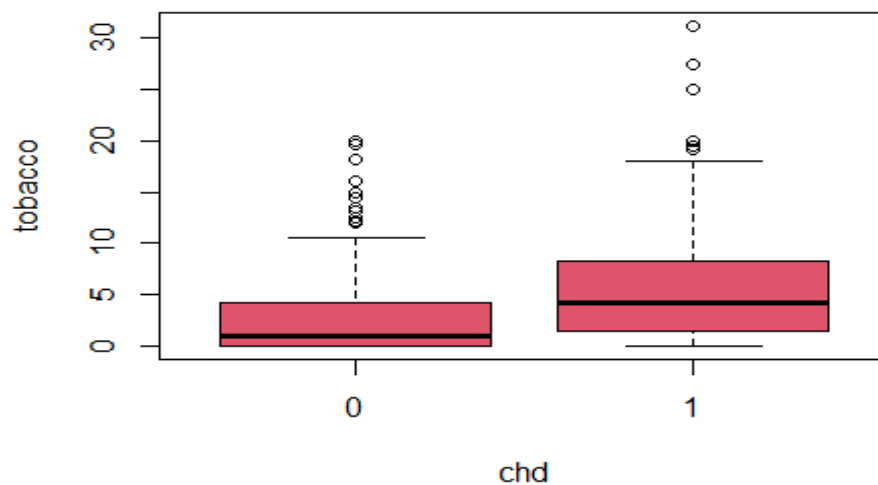
*#now we want to see the Box plot of each variable with our response:*

```
boxplot(sbp~chd,col=85)
```



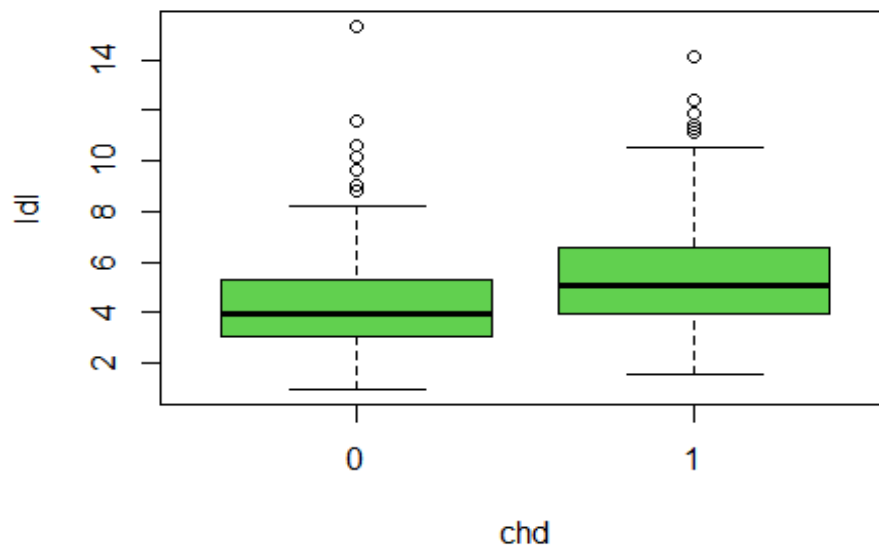
According to this Boxplot we can say that median of sbp for Having and not having coronary heart disease are equal but for more sbp values we have more coronary heart disease.

```
boxplot(tobacco~chd,col=2)
```



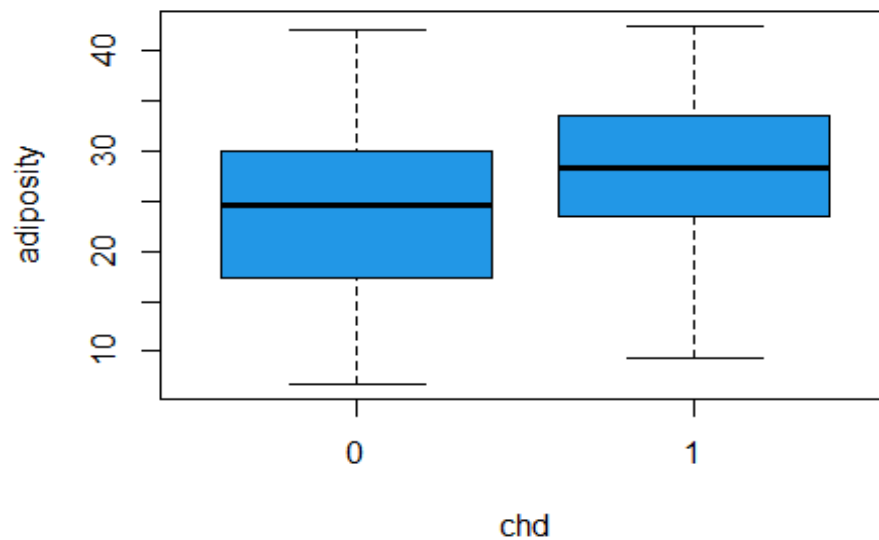
According to this Boxplot we can say that median of tobacco for Having and not having coronary heart disease are not equal and for more tobacco values we have more coronary heart disease.

```
boxplot(ldl~chd,col=3)
```



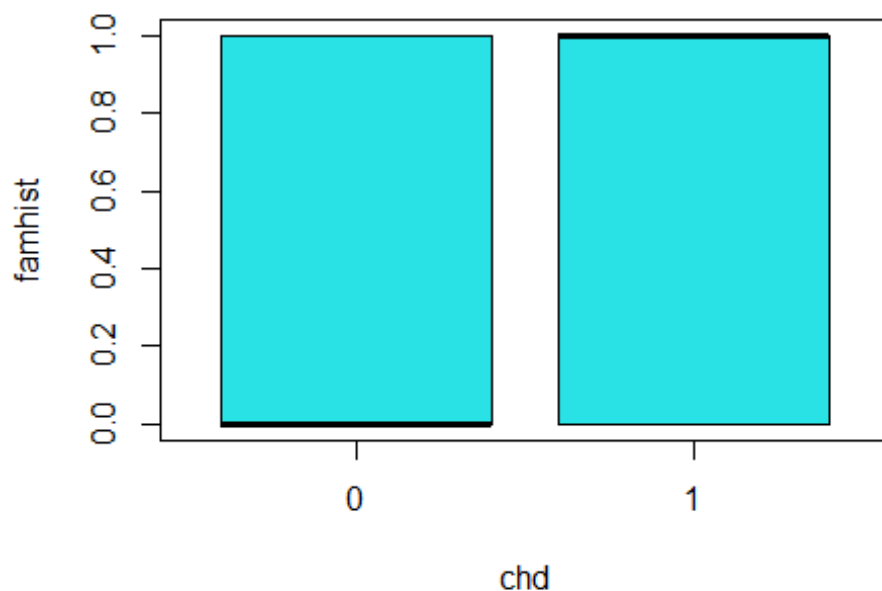
According to this Boxplot we can say that median of ldl for Having and not having coronary heart disease are not equal and for more ldl values we have more coronary heart disease.

```
boxplot(adiposity~chd,col=4)
```



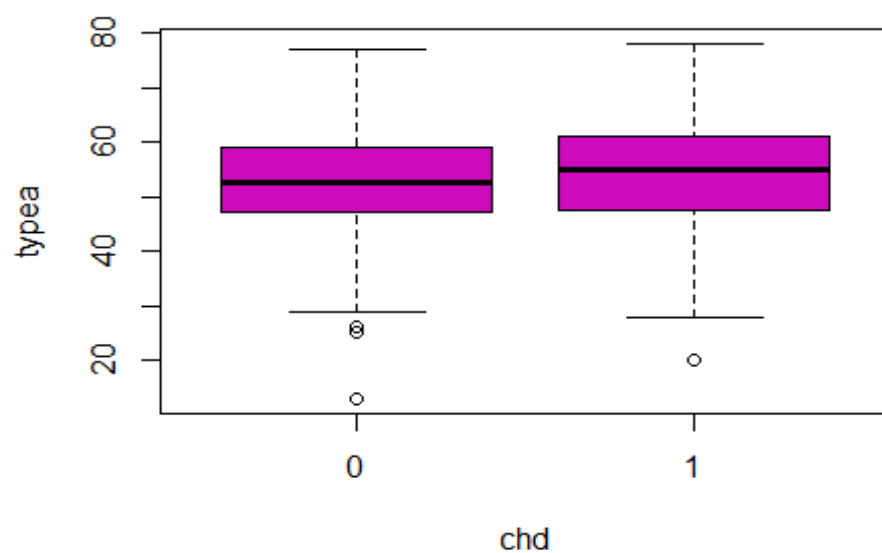
According to this Boxplot we can say that median of adiposity for Having and not having coronary heart disease are not equal and for adiposity values between (0,25) we can say that we don't have coronary heart disease and for more adiposity values we have more and more coronary heart disease.

```
boxplot(famhist~chd,col=5)
```



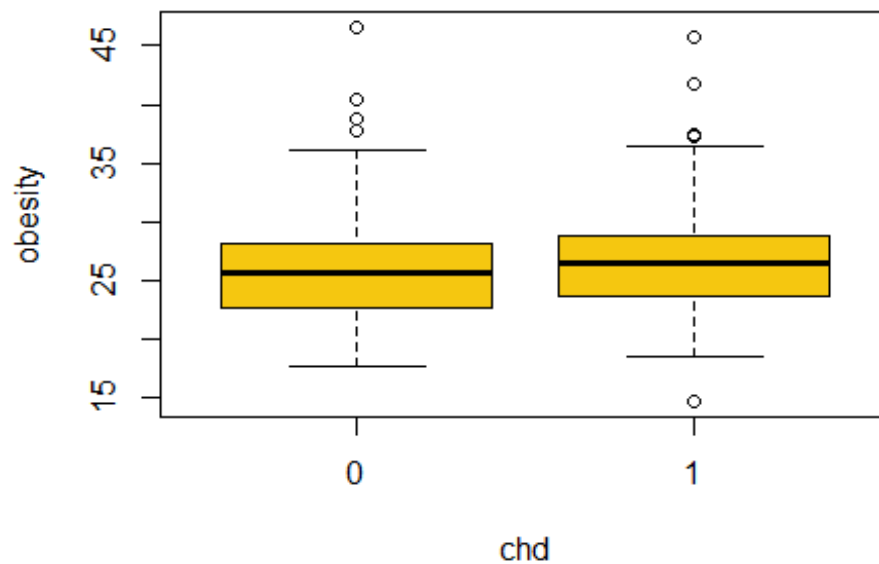
According to this Boxplot we can say that median of famhist for Having and not having coronary heart disease are not equal but for more famhist values we have more coronary heart disease.

```
boxplot(typea~chd,col=6)
```



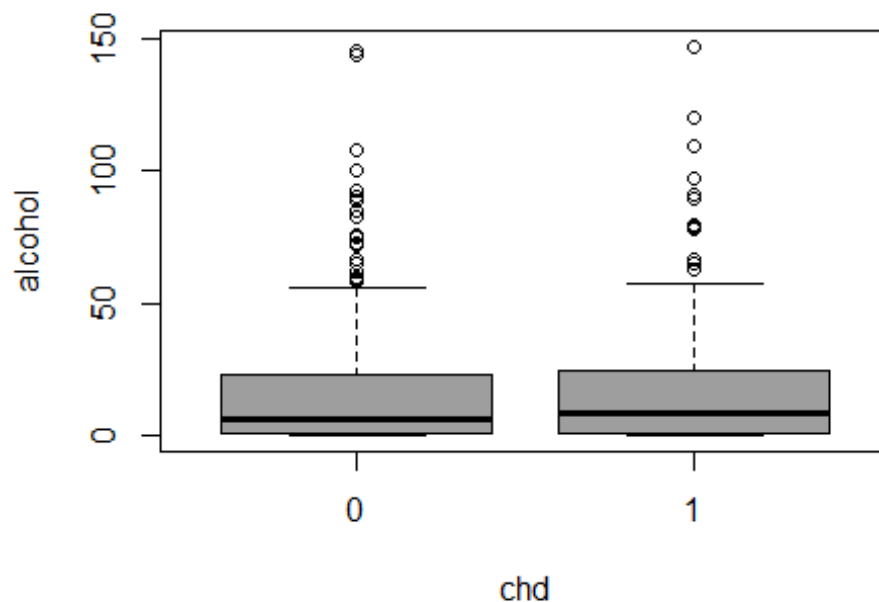
According to this Boxplot we can say that median of typea for Having and not having coronary heart disease are equal.

```
boxplot(obesity~chd,col=7)
```



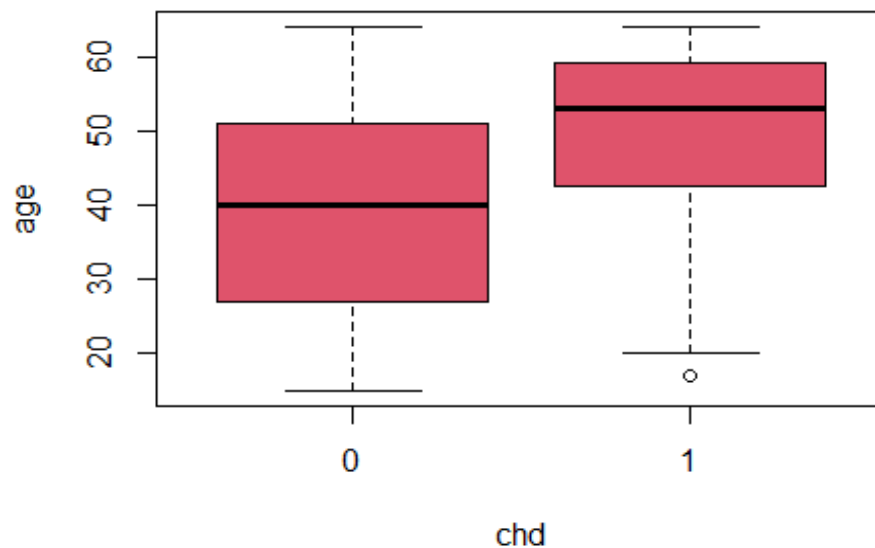
According to this Boxplot we can say that median of obesity for Having and not having coronary heart disease are equal.

```
boxplot(alcohol~chd,col=8)
```



According to this Boxplot we can say that median of alcohol for Having and not having coronary heart disease are equal.

```
boxplot(age~chd,col=10)
```



According to this Boxplot we can say that median of age for Having and not having coronary heart disease are not equal and for age values between (1,40) we can say that we dont have coronary heart disease and for more adiposity vlaues we have more and more coronary heart disease.

## Logstics Regression with severan variables and univariabes

*Logstics regression for each variable and response:*

```
fit1<-glm(chd~sbp,family = binomial)
coef(fit1)

## (Intercept)          sbp
## -3.35271610  0.01950936

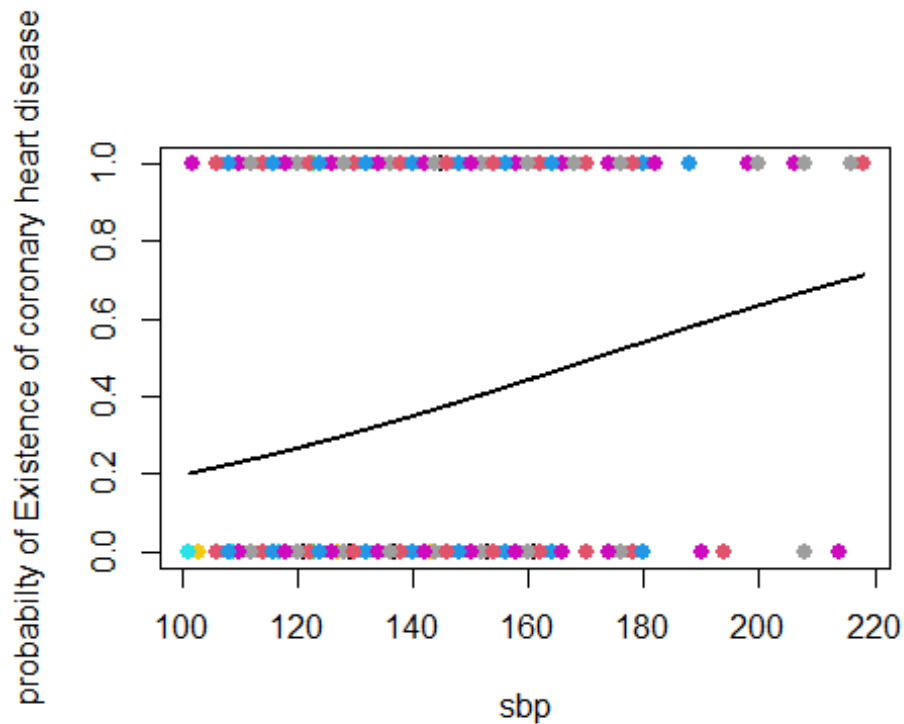
summary(fit1)

##
## Call:
## glm(formula = chd ~ sbp, family = binomial)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5405  -0.8982  -0.8009   1.3113   1.7836
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.352716   0.687698  -4.875 1.09e-06 ***
## sbp          0.019509   0.004863   4.012 6.02e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 596.11  on 461  degrees of freedom
## Residual deviance: 579.32  on 460  degrees of freedom
## AIC: 583.32
##
## Number of Fisher Scoring iterations: 4
```

According to the summary and coef function outputs we can say that we have posetive relationship (betha sbp = 0.019509) and our p-value for signifacting H0 (Betha0 = Betha1 = 0)is equal to 6.02e-05 and its less than alpha(0.05) so we say the H0 reject and the regression is signifact.



```
seq1<-data.frame(sbp=seq(min(sbp),max(sbp),len=10^4))
seq1$chd=predict(fit1,newdata= seq1,type="response")
plot(sbp,chd,col=seq1,pch=19,cex=1.1,ylab="probabilty of Existence of coronary
heart disease")
lines(chd~sbp , seq1 ,col=1,lwd=2)
```



According to top plot we can see the probabilty of Existence of coronary heart disease for sbp large or big values its more and more and we have positive relationship for example sbp is equal 220 our probabilty is equal 0.8.

```

fit2<-glm(chd~tobacco,family = binomial)
coef(fit2)

## (Intercept)      tobacco
## -1.1894300    0.1452696

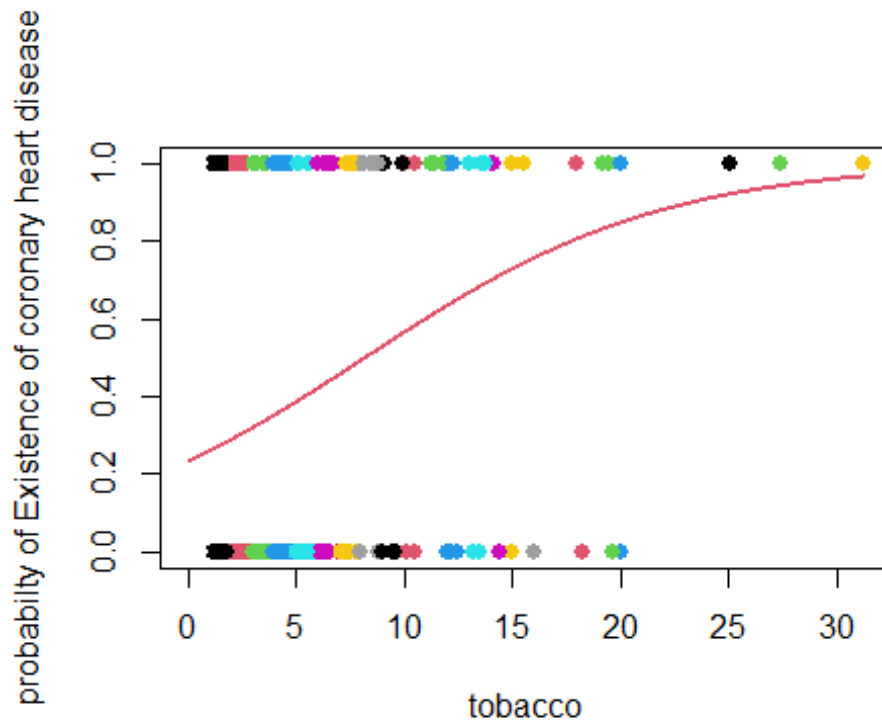
summary(fit2)

##
## Call:
## glm(formula = chd ~ tobacco, family = binomial)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.9397  -0.8467  -0.7290   1.1997   1.7060
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.18943    0.13900  -8.557  < 2e-16 ***
## tobacco      0.14527    0.02476   5.866 4.46e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 596.11  on 461  degrees of freedom
## Residual deviance: 554.65  on 460  degrees of freedom
## AIC: 558.65
##
## Number of Fisher Scoring iterations: 4

```

According to the summary and coef function outputs we can say that we have positive relationship (beta tobacco = 0.14527) and our p-value for signifacting H0 (Beta0 = Beta1 = 0) is equal to 4.46e-09 and its less than alpha(0.05) so we say the H0 reject and the regression is signifact.

```
seq2<-data.frame(tobacco=seq(min(tobacco),max(tobacco),len=10^4))
seq2$chd=predict(fit2,newdata= seq2,type="response")
plot(tobacco,chd,col=tobacco,pch=19,cex=1.1,ylab="probabilty of Existence of
coronary heart disease")
lines(chd~tobacco , seq2 ,col=2,lwd=2)
```



According to top plot we can see the probabilty of Existence of coronary heart disease for tobacco large or big its more and more values and we have posetive relationship , for example tobacco is more and equal 25 our probabiltiy is more than 0.95.

```

fit3<-glm(chd~ldl,family = binomial)
coef(fit3)

## (Intercept)          ldl
## -1.9686681    0.2746613

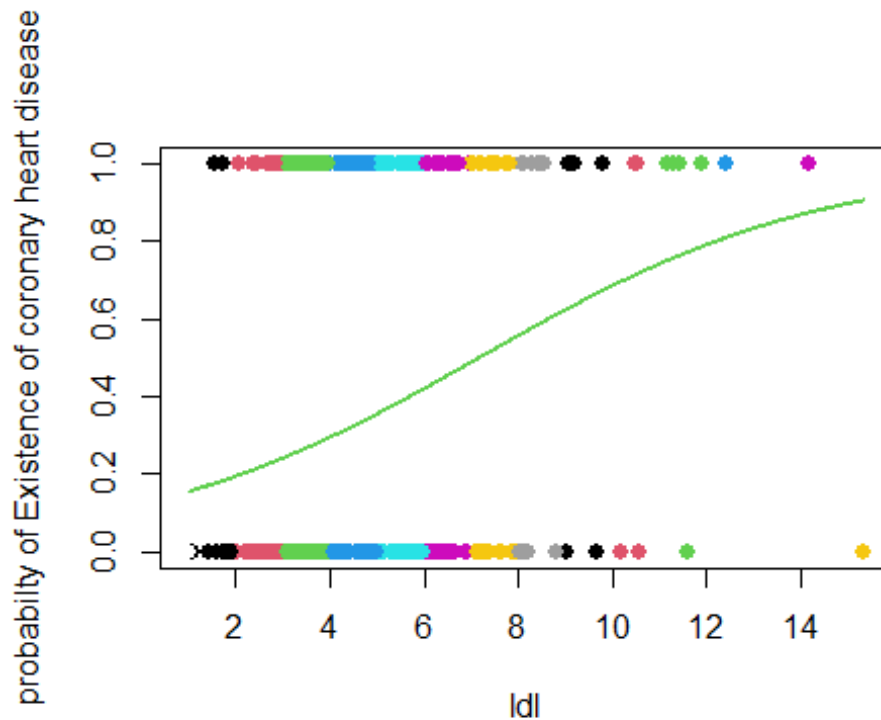
summary(fit3)

##
## Call:
## glm(formula = chd ~ ldl, family = binomial)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1647  -0.8948  -0.7426   1.2688   1.8637
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.96867    0.27308  -7.209 5.63e-13 ***
## ldl          0.27466    0.05164   5.319 1.04e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 596.11  on 461  degrees of freedom
## Residual deviance: 564.28  on 460  degrees of freedom
## AIC: 568.28
##
## Number of Fisher Scoring iterations: 4

```

According to the summary and coef function outputs we can say that we have positive relationship (beta tobacco = 0.27466) and our p-value for signifacting H0 (Beta0 = Beta1 = 0) is equal to 1.04e-07 and its less than alpha(0.05) so we say the H0 reject and the regression is signifact.

```
seq3<-data.frame(ldl=seq(min(ldl),max(ldl),len=10^4))
seq3$chd=predict(fit3,newdata= seq3,type="response")
plot(ldl,chd,col=ldl,pch=19,cex=1.1,ylab="probabilty of Existence of coronary
heart disease")
lines(chd~ldl, seq3 ,col=3,lwd=2)
```



According to top plot we can see the probabilty of Existence of coronary heart disease for ldl large or big values its more and more and we have posetive relationship, for example tobacco is more and equal 12 our probabily is more than 0.8.

```

fit4<-glm(chd~adiposity,family = binomial)
coef(fit4)

## (Intercept)    adiposity
## -2.56922635    0.07410225

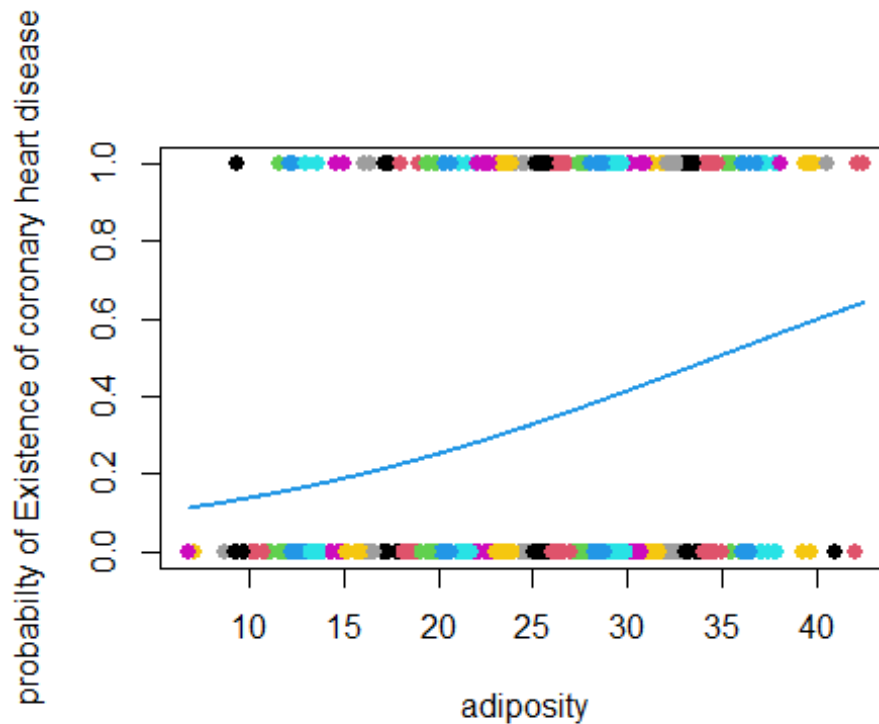
summary(fit4)

##
## Call:
## glm(formula = chd ~ adiposity, family = binomial)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4170  -0.9554  -0.6960   1.2228   2.0081
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.56923    0.38712  -6.637 3.21e-11 ***
## adiposity    0.07410    0.01396   5.308 1.11e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 596.11  on 461  degrees of freedom
## Residual deviance: 565.05  on 460  degrees of freedom
## AIC: 569.05
##
## Number of Fisher Scoring iterations: 4

```

According to the summary and coef function outputs we can say that we have positive relationship (beta adiposity = 0.07410) and our p-value for signifacting H0 (Beta0 = Beta1 = 0) is equal to 1.11e-07 and its less than alpha(0.05) so we say the H0 reject and the regression is signifact.

```
seq4<-data.frame(adiposity=seq(min(adiposity),max(adiposity),len=10^4))
seq4$chd=predict(fit4,newdata= seq4,type="response")
plot(adiposity,chd,col=adiposity,pch=19,cex=1.1,ylab="probabilty of Existence
of coronary heart disease")
lines(chd~adiposity, seq4 ,col=4,lwd=2)
```



According to top plot we can see the probabilty of Existence of coronary heart disease for adiposity large or big values its more and more and we have posetive relationship, for example adiposity is more and equal 35 our probabilty is more than 0.5.

```

fit5<-glm(chd~famhist,family = binomial)
coef(fit5)

## (Intercept)      famhist
##   -1.168993     1.168993

summary(fit5)

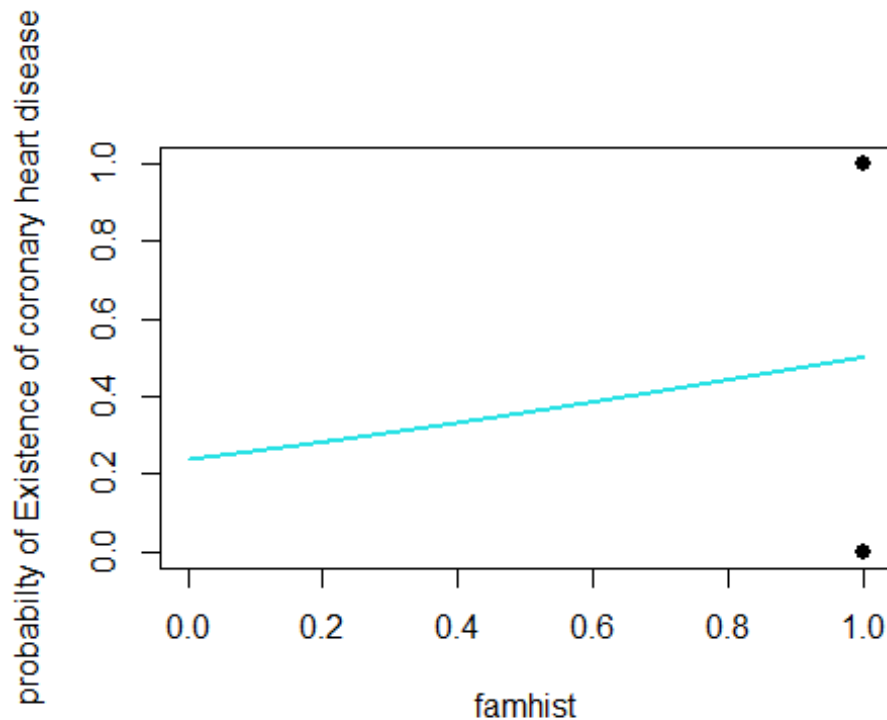
##
## Call:
## glm(formula = chd ~ famhist, family = binomial)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1774  -0.7356  -0.7356   1.1774   1.6968
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.1690     0.1431  -8.169 3.12e-16 ***
## famhist       1.1690     0.2033   5.751 8.85e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 596.11  on 461  degrees of freedom
## Residual deviance: 561.89  on 460  degrees of freedom
## AIC: 565.89
##
## Number of Fisher Scoring iterations: 4

```

According to the summary and coef function outputs we can say that we have positive relationship (beta famhist = 1.1690) and our p-value for signifacting H0 (Beta0 = Beta1 = 0) is equal to 8.85e-09 and its less than alpha(0.05) so we say the H0 reject and the regression is signifact.



```
seq5<-data.frame(famhist=seq(min(famhist),max(famhist),len=10^4))
seq5$chd=predict(fit5,newdata= seq5,type="response")
plot(famhist,chd,col=famhist,pch=19,cex=1.1,ylab="probabilty of Existence of
coronary heart disease")
lines(chd~famhist, seq5 ,col=5,lwd=2)
```



According to top plot we can see the probabilty of Existence of coronary heart disease for famhist large or big values its more and more and we have posetive relationship. but i think that its not good variable for our multiple logsticks regression.

```

fit6<-glm(chd~typea,family = binomial)
coef(fit6)

## (Intercept)      typea
## -1.84469292  0.02263029

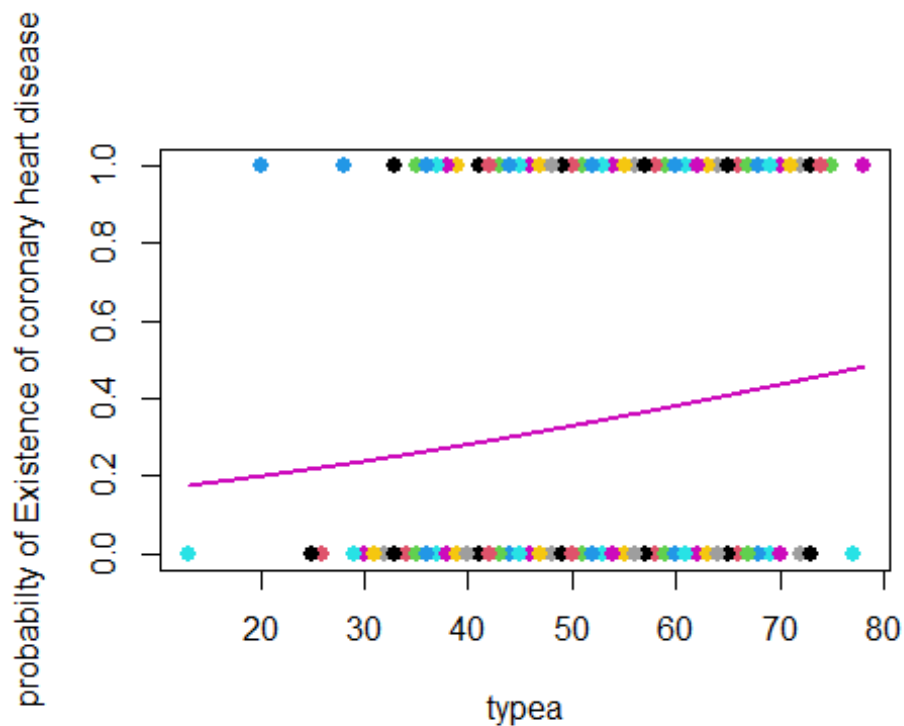
summary(fit6)

##
## Call:
## glm(formula = chd ~ typea, family = binomial)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1343  -0.9440  -0.8602   1.3874   1.7967
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.84469    0.56068  -3.290  0.0010 **
## typea        0.02263    0.01026   2.205  0.0275 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 596.11  on 461  degrees of freedom
## Residual deviance: 591.12  on 460  degrees of freedom
## AIC: 595.12
##
## Number of Fisher Scoring iterations: 4

```

According to the summary and coef function outputs we can say that we have weak positive relationship (beta typea = 0.02263) and our p-value for signifacting H0 (Beta0 = Beta1 = 0) is equal to 0.0275 and its less than alpha(0.05) so we say the H0 reject and the regression is signifact.

```
seq6<-data.frame(typea=seq(min(typea),max(typea),len=10^4))
seq6$chd=predict(fit6,newdata= seq6,type="response")
plot(typea,chd,col=typea,pch=19,cex=1.1,ylab="probabilty of Existence of
coronary heart disease")
lines(chd~typea, seq6 ,col=6,lwd=2)
```



According to top plot we can see the probabilty of Existence of coronary heart disease for typea large or big values its more and more and we have posetive relationship . but i think that its not good variable for our multiple logsticks regression.

```

fit7<-glm(chd~obesity,family = binomial)
coef(fit7)

## (Intercept)      obesity
## -1.92831227  0.04941705

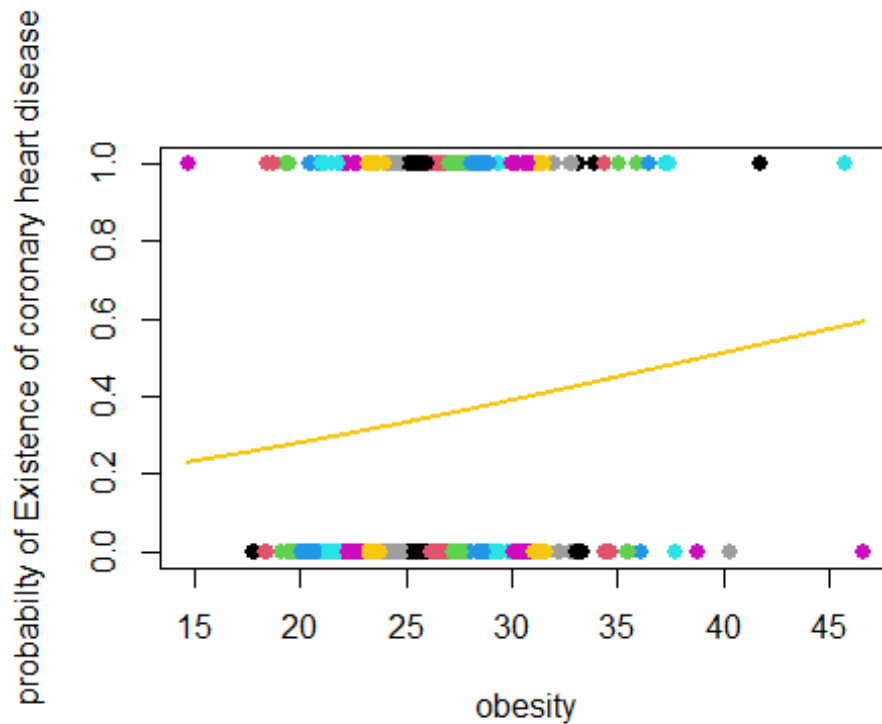
summary(fit7)

##
## Call:
## glm(formula = chd ~ obesity, family = binomial)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3396  -0.9257  -0.8558   1.4021   1.7116
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.92831    0.61692  -3.126  0.00177 **
## obesity      0.04942    0.02318   2.132  0.03302 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 596.11  on 461  degrees of freedom
## Residual deviance: 591.53  on 460  degrees of freedom
## AIC: 595.53
##
## Number of Fisher Scoring iterations: 4

```

According to the summary and coef function outputs we can say that we have weak positive relationship (beta obesity = 0.04942) and our p-value for signifacting H0 (Beta0 = Beta1 = 0) is equal to 0.03302 and its less than alpha(0.05) so we say the H0 reject and the regression is signifact.

```
seq7<-data.frame(obesity=seq(min(obesity),max(obesity),len=10^4))
seq7$chd=predict(fit7,newdata= seq7,type="response")
plot(obesity,chd,col=obesity,pch=19,cex=1.1,ylab="probabilty of Existence of
coronary heart disease")
lines(chd~obesity, seq7 ,col=7,lwd=2)
```



According to top plot we can see the probability of Existence of coronary heart disease for obesity large or big values its more and more and we have posetive relationship. but i think that its not good variable for our multiple logsticks regression.

```

fit8<-glm(chd~alcohol,family = binomial)
coef(fit8)

## (Intercept)      alcohol
## -0.726007546  0.005197845

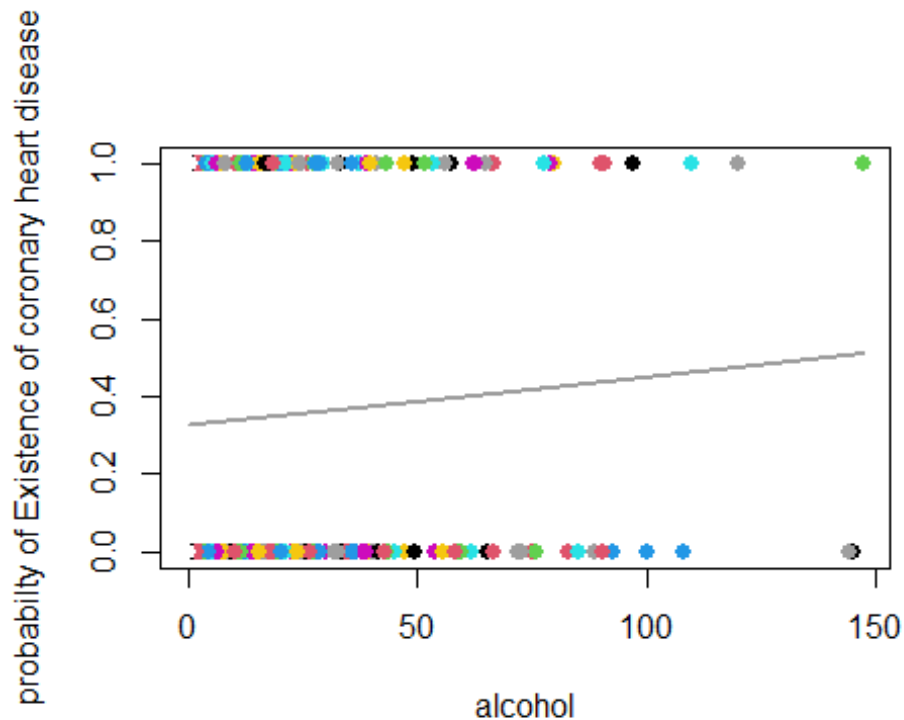
summary(fit8)

##
## Call:
## glm(formula = chd ~ alcohol, family = binomial)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1898  -0.9104  -0.8884   1.4415   1.4971
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.726008   0.120004  -6.050 1.45e-09 ***
## alcohol      0.005198   0.003892   1.336  0.182
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 596.11  on 461  degrees of freedom
## Residual deviance: 594.35  on 460  degrees of freedom
## AIC: 598.35
##
## Number of Fisher Scoring iterations: 4

```

According to the summary and coef function outputs we can say that we have weak positive relationship (beta alcohol = 0.005198) and our p-value for signifacting H0 (Beta0 = Beta1 = 0) is equal to 0.182 and its more than alpha(0.05) so we say the H0 accept and the regression isnt signifact.

```
seq8<-data.frame(alcohol=seq(min(alcohol),max(alcohol),len=10^4))
seq8$chd=predict(fit8,newdata= seq8,type="response")
plot(alcohol,chd,col=alcohol,pch=19,cex=1.1,ylab="probabilty of Existence of
coronary heart disease")
lines(chd~alcohol, seq8 ,col=8,lwd=2)
```



According to top plot we can see the probabilty of Existence of coronary heart disease for alcohol large or big values its more and more and we have posetive relationship. but i think that its not good variable for our multiple logsticks regression.

```

fit9<-glm(chd~age,family = binomial)
coef(fit9)

## (Intercept)          age
## -3.52171034  0.06410803

summary(fit9)

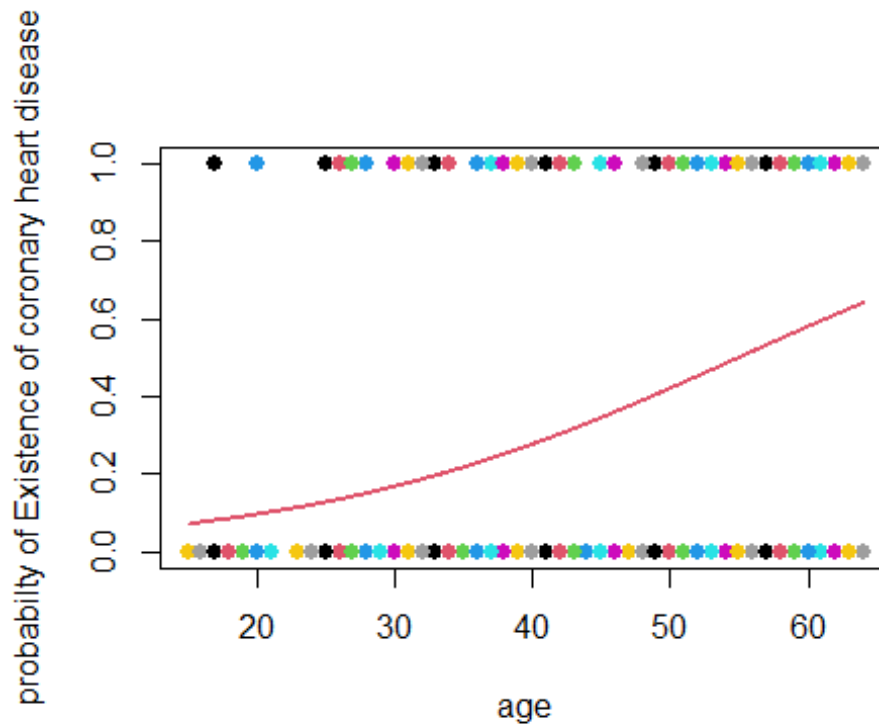
##
## Call:
## glm(formula = chd ~ age, family = binomial)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4321  -0.9215  -0.5392   1.0952   2.2433
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.521710   0.416031  -8.465  < 2e-16 ***
## age          0.064108   0.008532   7.513 5.76e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 596.11  on 461  degrees of freedom
## Residual deviance: 525.56  on 460  degrees of freedom
## AIC: 529.56
##
## Number of Fisher Scoring iterations: 4

```

According to the summary and coef function outputs we can say that we have weak positive relationship (beta age = 0.064108) and our p-value for signifacting H0 (Beta0 = Beta1 = 0) is equal to 5.76e-14 and its less than alpha(0.05) so we say the H0 reject and the regression is signifact.



```
seq9<-data.frame(age=seq(min(age),max(age),len=10^4))
seq9$chd=predict(fit9,newdata= seq9,type="response")
plot(age,chd,col=age,pch=19,cex=1.1,ylab="probabilty of Existence of coronary
heart disease")
lines(chd~age, seq9 ,col=10,lwd=2)
```



According to top plot we can see the probability of Existence of coronary heart disease for age large or big values its more and more and we have positive relationship , for example adiposity is more and equal 55 our probability is more than 0.5.

### *Logistics regression for all variable and response(Multiple logistic regression)*

```
full.fit<-glm(chd~.,data=Heart,family = binomial)
coef(full.fit)

##      (Intercept)      sbp      tobacco      ldl      adiposity
## -6.1507208650  0.0065040171  0.0793764457  0.1739238981  0.0185865682
##      famhist      typea      obesity      alcohol      age
##  0.9253704194  0.0395950250 -0.0629098693  0.0001216624  0.0452253496

summary(full.fit)

##
## Call:
## glm(formula = chd ~ ., family = binomial, data = Heart)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7781  -0.8213  -0.4387   0.8889   2.5435
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -6.1507209  1.3082600  -4.701 2.58e-06 ***
## sbp          0.0065040  0.0057304   1.135 0.256374
## tobacco      0.0793764  0.0266028   2.984 0.002847 **
## ldl          0.1739239  0.0596617   2.915 0.003555 **
## adiposity    0.0185866  0.0292894   0.635 0.525700
## famhist      0.9253704  0.2278940   4.061 4.90e-05 ***
## typea        0.0395950  0.0123202   3.214 0.001310 **
## obesity      -0.0629099  0.0442477  -1.422 0.155095
## alcohol       0.0001217  0.0044832   0.027 0.978350
## age          0.0452253  0.0121298   3.728 0.000193 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 596.11  on 461  degrees of freedom
## Residual deviance: 472.14  on 452  degrees of freedom
## AIC: 492.14
##
## Number of Fisher Scoring iterations: 5
```

According to this models we can say that the (sbp,adiposity,obesity,alcohol) variables p-values is more than 0.05 and they are not good for our model and we should remove them, the others variable have positive relationships with response except obesity.

### Reduce logstics model

```
Reduce.fit<-glm(chd~tobacco+ldl+famhist+typea+age,data=Heart,family =
binomial)
coef(Reduce.fit)

## (Intercept)      tobacco          ldl      famhist      typea      age
## -6.44644451  0.08037533  0.16199164  0.90817526  0.03711521  0.05046038

summary(Reduce.fit)

##
## Call:
## glm(formula = chd ~ tobacco + ldl + famhist + typea + age, family =
binomial,
##      data = Heart)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.9165  -0.8054  -0.4430   0.9329   2.6139
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -6.44644    0.92087  -7.000 2.55e-12 ***
## tobacco      0.08038    0.02588   3.106 0.00190 **
## ldl          0.16199    0.05497   2.947 0.00321 **
## famhist      0.90818    0.22576   4.023 5.75e-05 ***
## typea        0.03712    0.01217   3.051 0.00228 **
## age          0.05046    0.01021   4.944 7.65e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 596.11  on 461  degrees of freedom
## Residual deviance: 475.69  on 456  degrees of freedom
## AIC: 487.69
##
## Number of Fisher Scoring iterations: 5
```

We can see that the reduce models outputs show that the (sbp,adiposity,obesity,alcohol) variable were not important and dont have effects on response.

the Reduce model is best model without any bad variable and all of them are signifant and we have good predict for our response,

we can see the logstics predicton function here for reduce model:

y = chd , x1 = tobacco , x2 = ldl , x3 = famhist , x4 = typea , x5= age.

$$P(X) = P(y = 1|X) = \frac{\exp\{\beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_5x_5\}}{1 + \exp\{\beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_5x_5\}}$$
$$= \frac{\exp\{-6.44644 + (0.08038 * x_1) + (0.16199 * x_2) + (0.90818 * x_3) + (0.03712 * x_4) + (0.05046 * x_5)\}}{1 + \exp\{-6.44644 + (0.08038 * x_1) + (0.16199 * x_2) + (0.90818 * x_3) + (0.03712 * x_4) + (0.05046 * x_5)\}}$$

End.