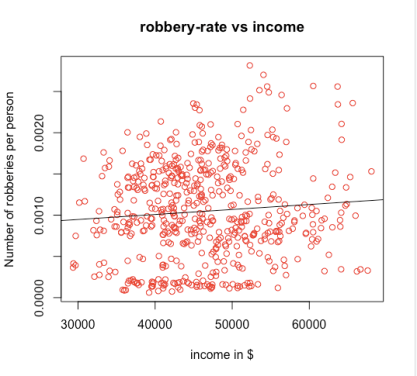
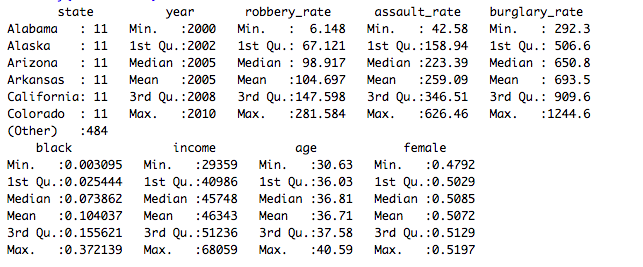
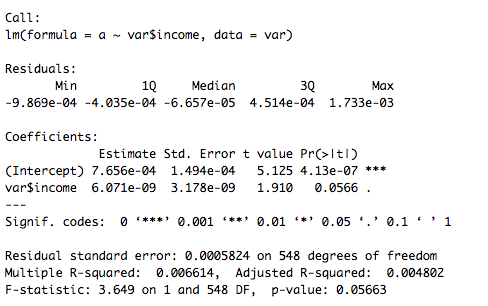
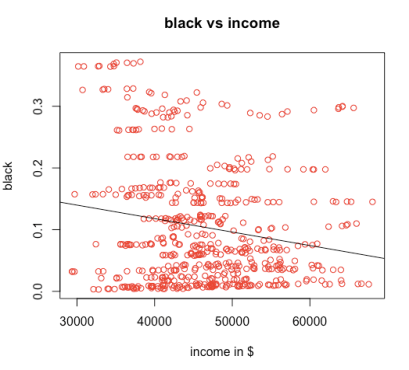
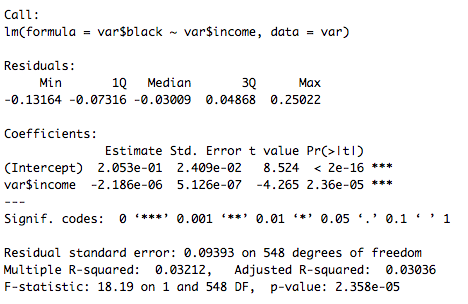
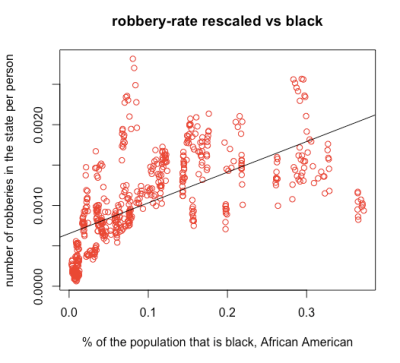
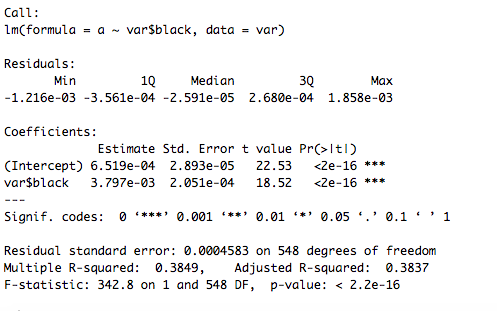
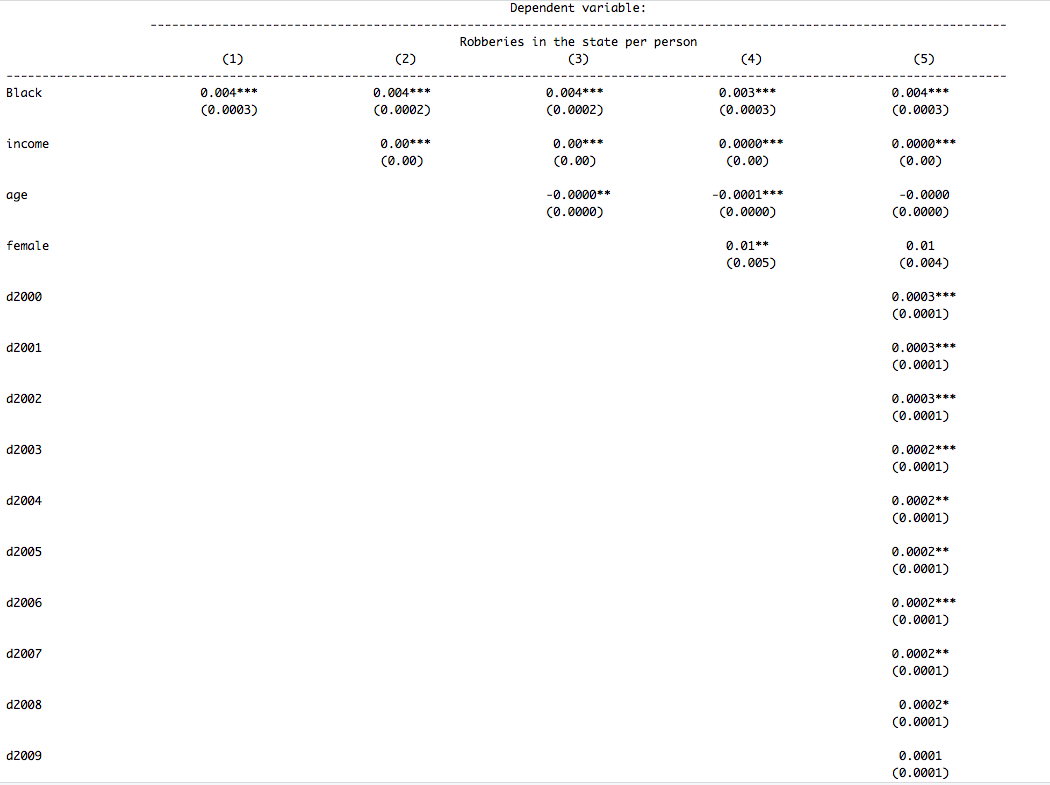
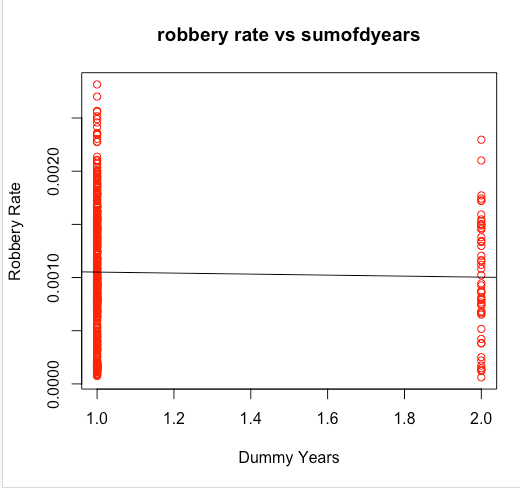
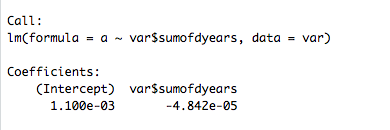
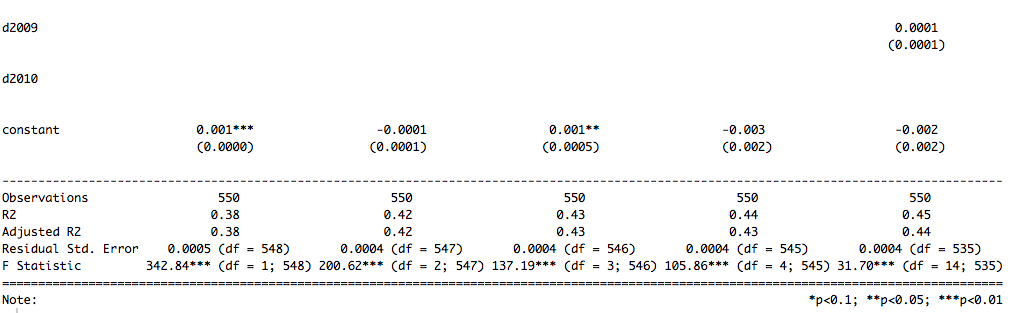
**Assignment 2 ECOM20001 (FINAL EDIT)  
Michael Le (998211)  
  
Question 1**In the US alone, the average assault rate 259.09 which is about two times the  
robbery rate of 104.697 and 3.2 times the rate of burglary rate 693.5 in the state per 100,000 people. Given the average age around 37 years, 10.4% caused by African Americans with the average income of $46,343. (NOTE: For standard derivations see Appendix A1 Question 1) **Question 2**

****

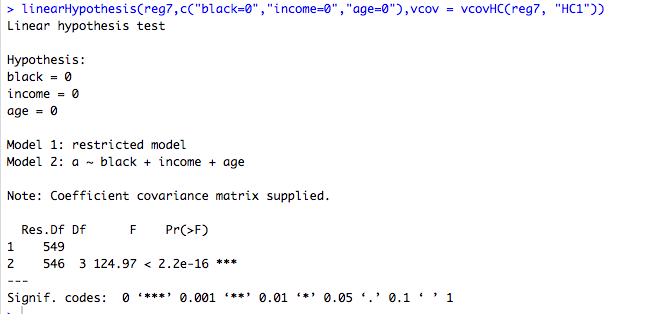
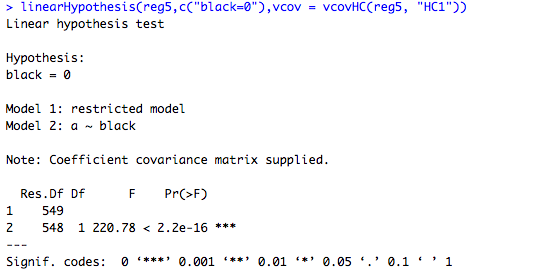
****

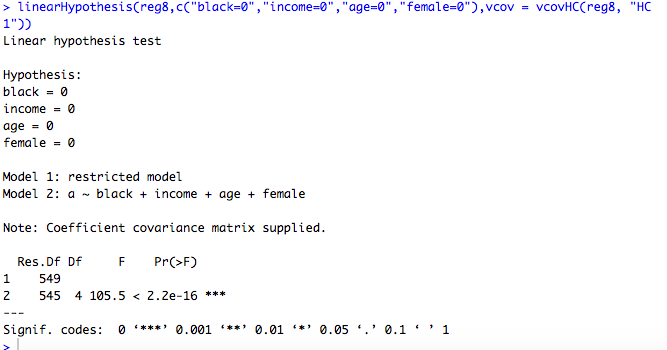
****

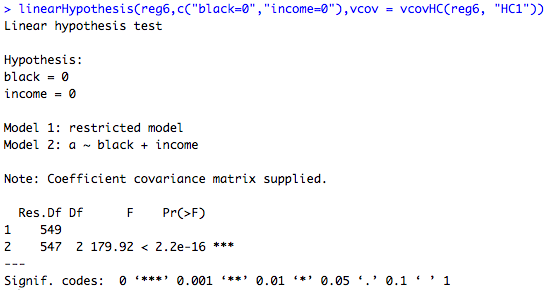
**Question 3**From Lecture 6 Slides 14 to 20!!!!  
If the repressor (black(i)) is correlated with a variable that has been omitted from the analysis (Income(I)) determines the robbery-rate (dependent). Across two different scatter plots the correlations for robbery rate (0.0813288) and black vs. income (-0.1792262) makes the sign of the OVB to become negative. There will be a negative bias in beta (1).hat to the true value of beta (1). Hence on the single linear regression model fails. Given the R^2 for robbery rate vs. black is 0.3837 which is relatively small amongst other independent variables that will be tested in later questions in this assignment. **Question 4  
**There are vertical points at 1 and 2 for the dummy Year Variables as Robbery  
rate increases because there only values between 0 and 1 and adding all the  
years from 2000 to 2010.This can be avoided by simply adding more independent (regressors) to var$sumofdyears into the regression model. In which we refer from Tutorial 7 or 8 to use the stargazer() and AER packages to help us determine outputs for well-formatted regression tables, with multiple models side-by-side, as well as for  
summary statistics tables, data frames, vectors and matrices. **Question 5**

****

**Regression F-Statistic, accounting for heteroskadasticity**

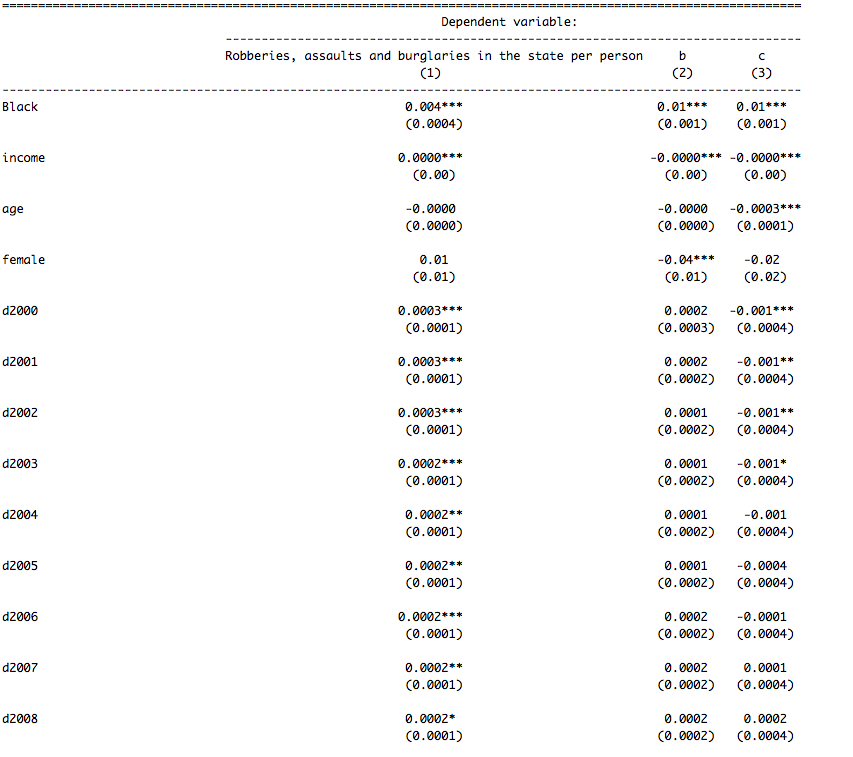
**CASE 1 reg5 CASE 3 reg7**  
****

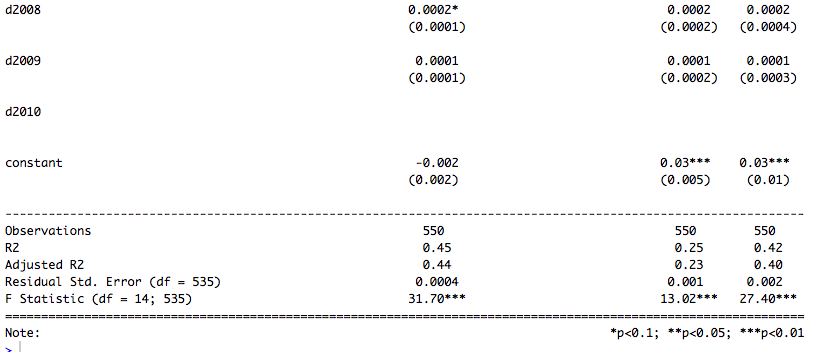
**CASE 2 reg6 CASE 4 reg8**



**NOTE:** 'stargazer' reports homoskedastic regression F-statistics  
to compute the correct F-statistic for each regression, we can use the linearHypothesis()  
command which allows us to compute joint hypothesis tests, which includes the  
regression F-statistics (which tests the null that all the coefficients are jointly 0).  
reg5 Regression F-statistic, accounting for heteroskedasticity.  
 **Question 6  
Part a)** There is no change on the coefficent in columns (1) and (2) compare to Q2 and Q3**.  
   
  
Part b)**At Column (4) the coefficent starts to settle down as (-0.0001 the coefficent is statistically insufficent) as 1 per cent of females increases. In Reg(5) the coefficent of black is signifant at (p<0.01) not (p< 0.05) which is much lower.  
 **Part c)**The base group in Column (5), that did not include in this was d2010 given no certain value. While we may discard d2001 to d2002, and d2004 to d2008 except d2009 with the coefficient 0.0001. Given each of consecutive coefficients 0.0003, 0.0002 respectively. Given the coefficient of 0.004 for the robbery rate is statistically significant but when we add the female and the dummy years (excluding the base group) it is statistically significantly lower given the p value less than 0.01.Which all coefficients are non zero and positive except d2010 that has no  
value.

**Part d)** (1) Interpreting a 1-unit change in black which means there is an increase of 0.4% in robbery rate.  
 (2) Interpreting by 1 standard derivation (given that the standard error(0.0004) = standard derivation/ square root of n (550)) gives us 9.38 x 10^-3 counts as 1 standard derivation times the mean(robbery per person) predicted increase the robbery rate for every African American. In conclusion we do not know exactly what the mean is in (2) given there are many residuals to work with (income, female, age and black). (1) is more relevent than (2). **Part e)**It would be more useful to present the predicted changes in robbery rates for a one-standard deviation change in black than a one-unit change in black because the former is a “standard” change in black in the data, whereas the latter is an extreme 0 to 1 change from the theoretical min to the max in the data, which is virtually impossible in reality. Despite how small these changes are made depending on every citizen regardless  
of their ethnicities.

**Question 7**

****

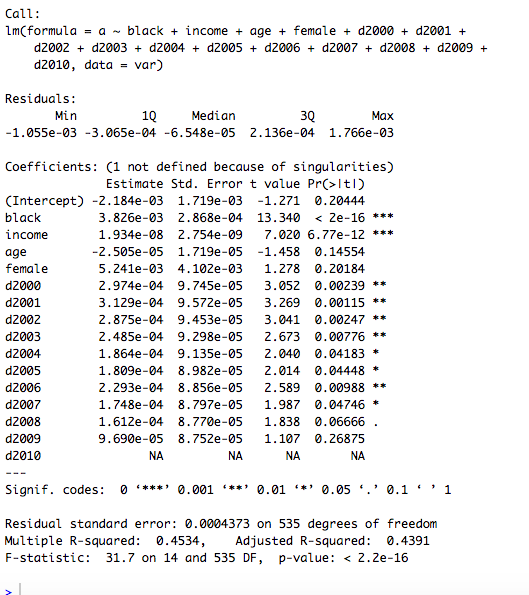
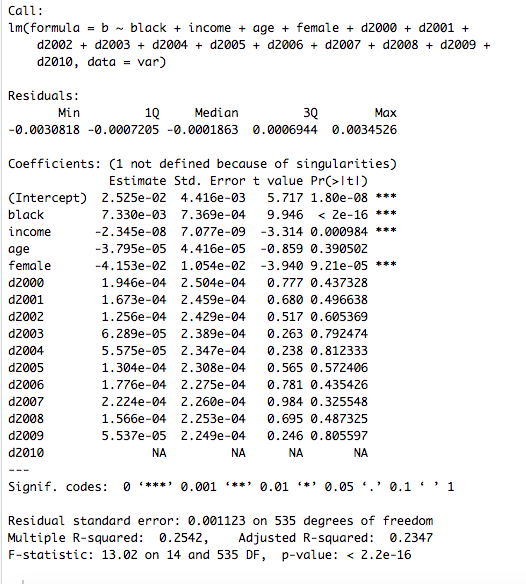
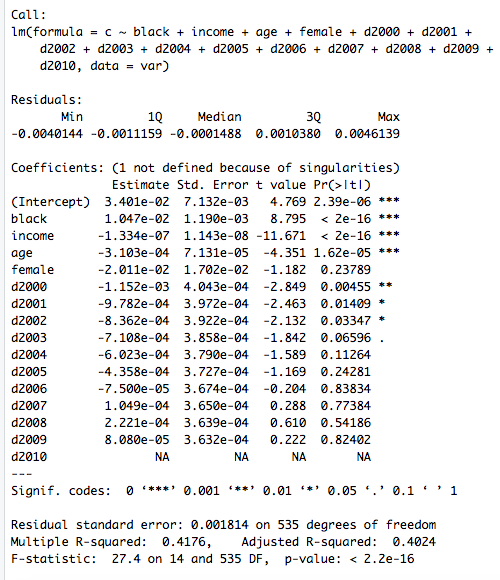
**Question 8  
Part a)** Using the 5% significance level, both coefficients in Columns (2) and (3) are  
significantly lower than 5%. 1 unit change increases 1% of burglary and assault rates  
for every African American.  **Part b)**By definition the best predicted is Column (1) with (3) slightly behind by 0.04 in difference in R^2 adjusted to  
make it slighly accurate predictions. **Part c)**

**Case 1: Robbery Rate vs. Black**[0.003263872, 0.004388128] 95% CI

**Case 2: Assault rate vs. Black**

[0.005885676,0.008774324] 95% CI

**Case 3: Burglary rate vs. Black**

[0.0081376, 0.0128024] 95% CI **Part d)   
Figure 1 Figure 2 Figure 3**

**-** F-statistic: 31.7 - F-Statistic: 13.02 - F-Statistic: 27.4

- p-value: 2.2x10^-16 - p-value: 2.2x10^-16 - p-value: 2.2x10^-16

- dof: 535 -dof: 535 - dof: 535

FIgure 1's given R^2 squared is bigger than Figure 2 and 3 values depending on the crimes were committed in the US independent to black, age, gender and income. All p-values (each of them are statistically significantly lower) with the same value and the same degrees of freedom. Despite this, data shows that there are more crimes in robbery compared to burglary and assault. This could likely due that they are disadvantaged and stereotype amongst other ethnicities among US society. This can be prevented by serving longer and extended sentences in prisons can help prevents crime records, more robberies and reports increasing from the FBI and Bureau of Justice Statistics. Decreases the crime rate coefficients in robbery, burglary and assault among the population that is black and increase their incomes (saving their incomes) and avoid heavy penalties, police checking over time to further decrease even further.

**Appendix A1 (Question 1)**

var = read.csv("as2\_crime.csv")  
summary(var)  
  
*Where interested in the actual information (i.e ignore the year, state)*  
sd(var$robbery\_rate) #58.38495  
sd(var$burglary\_rate) #234.6821  
sd(var$assault\_rate) #128.4194  
sd(var$black) #0.09538564  
sd(var$income) #7820.835  
sd(var$age) #1.533901  
sd(var$female) #0.007407594

*To rescale, IN THIS CASE WE ARE INTERESTED ONLY IN THE CRIME RATES ONLY!!!*  
a = var$robbery\_rate/10^5  
b = var$assault\_rate/10^5  
c = var$burglary\_rate/10^5  
  
var\_rescale = data.frame(a,b,c)  
var #in 100,000's

**Appendix A2 (Question 2)**

#Case 1: robbery\_rate vs. black  
reg1=lm(a~var$black,data=var)  
summary(reg1)cor(var$black,a) # 0.6203645  
plot(var$black,a,  
     main="robbery-rate rescaled vs black",  
     xlab="% of the population that is black, African American",  
     ylab="number of robberies in the state per person",  
     col="red",  
     pch=1)  
abline(reg1)  
  
  
#Case 2: robbery\_rate vs. income  
reg2=lm(a~var$income,data=var)  
summary(reg2)  
  
cor(var$income,a) #0.0813288  
plot(var$income,a,  
     main="robbery-rate vs income",  
     xlab="income in $",  
     ylab="Number of robberies per person",  
     col="red",  
     pch=1)  
abline(reg2)

#Case 3: black vs. income  
reg3=lm(var$black~var$income,data=var)  
summary(reg3)  
  
cor(var$income,var$black) #-0.1792262  
plot(var$income,var$black,main="black vs income",xlab="income in $",ylab="black",col="red",pch=1)  
  
abline(reg3)

**Appendix A3 (Question 4)**

We denote these dummy variables first in vector form  
var$d2000 = as.numeric(var$year == 2000)  
var$d2001 = as.numeric(var$year == 2001)  
var$d2002 = as.numeric(var$year == 2002)  
var$d2003 = as.numeric(var$year == 2003)  
var$d2004 = as.numeric(var$year == 2004)  
var$d2005 = as.numeric(var$year == 2005)  
var$d2006 = as.numeric(var$year == 2006)  
var$d2007 = as.numeric(var$year == 2007)  
var$d2008 = as.numeric(var$year == 2008)  
var$d2009 = as.numeric(var$year == 2009)  
var$d2010 = as.numeric(var$year == 2010)  
var$sumofdyears = var$d2000 + var$d2001 + var$d2002 + var$d2003 + var$d2004 + var$d2004 + var$d2005 + var$d2006 + var$d2007 + var$d2008 + var$d2009 + var$d2010    
  
reg4 = lm(a~var$sumofdyears, data = var)  
plot(var$sumofdyears,a,  
     main="robbery rate vs sumofdyears",  
     xlab="Dummy Years",  
     ylab="Robbery Rate",  
     col="red",  
     pch=1)  
abline(reg4)

**Appendix A4 (Question 5)**

*Refer from Tutorial 7 or 8  
  
Summarize the constant variable - it's always 1!*  
summary(var$sumofdyears)  
  
 *Regressions, storing and outputting the regression results using stargazer()*  
 See tute7.pdf and tute7.R for a detailed introduction to stargazer()  
  
#Case 1 (Robbery\_rate vs. black)  
reg5=lm(a~black,data=var)  # Regression estimates for reg1  
cov1=vcovHC(reg5, type = "HC1")     # The next 2 lines produce heteroskedasticity-robust standard errors for reg1  
se1=sqrt(diag(cov1))  
  
#Case 2 (Robbery\_rate vs. black+income)  
reg6 = lm(a~black+income,data=var)  
cov2 = vcovHC(reg6, type = "HC1")  
se2 = sqrt(diag(cov2))  
  
#Case 3 (Robbery\_rate vs. black+income+age)  
reg7 = lm(a~black+income+age,data=var)  
cov3 = vcovHC(reg7, type = "HC1")  
se3 = sqrt(diag(cov3))  
  
#Case 4 (Robbery\_rate vs. black+income+age+female)  
reg8 = lm(a~black+income+age+female,data=var)  
cov4= vcovHC(reg8, type = "HC1")  
se4 = sqrt(diag(cov4))  
  
#Case 5 (Robbery\_rate vs. black+income+age+female+sumofdyears)  
  
reg9 = lm(a~black+income+age+female+ d2000 + d2001 + d2002 + d2003 + d2004 + d2005 + d2006 + d2007 + d2008 + d2009 + d2010,data=var)  
cov5 = vcovHC(reg9, type = "HC1")  
se5 = sqrt(diag(cov5))  
  
Regression output table for the 5 regressions reg 5,reg6,reg7,reg8,reg9  
stargazer(reg5, reg6, reg7, reg8, reg9, type="text",  
          se=list(se1, se2, se3, se4,se5),  
          digits=2,  
          dep.var.labels=c("Robberies in the state per person"),  
covariate.labels=c("Black","income","age","female","d2000","d2001","d2002","d2003","d2004","d2005","d2006","d2007","d2008","d2009","d2010","constant"),out="reg\_output.txt")

**Appendix A5 (Question 7)**

##Case 1: (Robbery\_rate vs. black+income+age+female+sumofdyears) from Case5 from Q5  
reg9 = lm(a~black+income+age+female+ d2000 + d2001 + d2002 + d2003 + d2004 + d2005 + d2006 + d2007 + d2008 + d2009 + d2010,data=var)  
summary(reg9)  
cov5 = vcovHC(reg9, type = "HC1")  
se5 = sqrt(diag(cov5))  
  
  
  
##Case 2: (assult\_rate vs black+income+age+female+sumofdyears)  
reg10 = lm(b~black+income+age+female+ d2000 + d2001 + d2002 + d2003 + d2004 + d2005 + d2006 + d2007 + d2008 + d2009 + d2010,data=var)  
cov6 = vcovHC(reg10, type = "HC1")  
se6 = sqrt(diag(cov6))  
  
##Case 3: (burglary\_rate vs black+income+age+female+sumofdyears)  
reg11 = lm(c~black+income+age+female+ d2000 + d2001 + d2002 + d2003 + d2004 + d2005 + d2006 + d2007 + d2008 + d2009 + d2010,data=var)  
cov7 = vcovHC(reg11, type = "HC1")  
se7 = sqrt(diag(cov7))  
  
## Regression output table for the 3 regressions reg9, reg10, reg11  
stargazer(reg9, reg10, reg11, type="text",  
          se=list(se5, se6,se7),  
          digits=2,  
          dep.var.labels=c("Robberies, assaults and burglaries in the state per person"),  
          covariate.labels= c("Black","income","age","female","d2000", "d2001", "d2002","d2003",  
        "d2004","d2005","d2006", "d2007", "d2008","d2009","d2010", "constant"),  
          out="reg\_output.txt")

**Appendix A6a (Question 8)**

(Recall from that Mutiple quiz 8, if it helps?)  
## **Case 1: Robbery Rate vs. Black**  
summary(reg9)  
mean1 = 0.003826  
zvalue1 = 1.96  
se1 = 0.0002868  
UB1 = mean1 + zvalue1\*se1  
LB1 = mean1 - zvalue1\*se1        
UB1  
LB1  ## [0.003263872, 0.004388128]  
  
## **Case 2: Assault rate vs. Black**  
summary(reg10)  
mean2 = 0.007330  
zvalue2 = 1.96  
se2 = 0.0007369  
UB2 = mean2 + zvalue2\*se2  
LB2 = mean2 - zvalue2\*se2 #[0.005885676,0.008774324]        
UB2  
LB2  
  
#**Case 3: Burglary rate vs. Black**  
summary(reg11)  
mean3 = 0.01047  
zvalue3 = 1.96  
se3 = 0.001190  
UB3 = mean3 + zvalue3\*se3  
LB3 = mean3 - zvalue3\*se3  #[0.0081376, 0.0128024]  
UB3  
LB3

**Appendix A6b (Question 8)**

Regression F-statistic, accounting for heteroskedasticity

##Case 1: Robbery- Rate:  
linearHypothesis(reg9,c("black\_scaled=0","income\_scaled=0",  
                        "age=0",  
                        "female=0"),vcov = vcovHC(reg9, "HC1")

Unable to use linearHypothesis, how to easily avoid a perfect multicollinearity.  
try summary(reg9)  
  
##Case 2: Assault- Rate (similar as before)  
summary(reg10)  
  
##Case 3: Burglary-Rate (similar to the previous two cases)  
summary(reg11)