

011852253 Assignment 5-a

SHARATH KUMAR KARNATI

2024-10-23

#Question 1:1)

(18 points) This question involves the use of multiple linear regression on the redwine(winequality-red.csv) data set available on Canvas in the Datasets for Assignments module. This is the same dataset used in Assignment 2. a. (6 points) Perform a multiple linear regression with pH as the response and all other variables except citric_acid as the predictors. Show a printout of the result (including coefficient, error, and t-values for each predictor). Comment on the output by answering the following questions: i) Which predictors appear to have a statistically significant relationship to the response? How do you determine this? ii) What does the coefficient for the free_sulfur_dioxide variable suggest, in simple terms?

```
library(dplyr)

redwine <- read.csv("/Users/sharathkarnati/Desktop/DS/Assignment 5/winequality-red.csv")

summary(redwine)
```

```
## fixed_acidity volatile_acidity citric_acid residual_sugar
## Min. : 4.60 Min. :0.1200 Min. :0.000 Min. : 0.900
## 1st Qu.: 7.10 1st Qu.:0.3900 1st Qu.:0.090 1st Qu.: 1.900
## Median : 7.90 Median :0.5200 Median :0.260 Median : 2.200
## Mean : 8.32 Mean :0.5278 Mean :0.271 Mean : 2.539
## 3rd Qu.: 9.20 3rd Qu.:0.6400 3rd Qu.:0.420 3rd Qu.: 2.600
## Max. :15.90 Max. :1.5800 Max. :1.000 Max. :15.500
## chlorides free_sulfur_dioxide total_sulfur_dioxide density
## Min. :0.01200 Min. : 1.00 Min. : 6.00 Min. :0.9901
## 1st Qu.:0.07000 1st Qu.: 7.00 1st Qu.: 22.00 1st Qu.:0.9956
## Median :0.07900 Median :14.00 Median : 38.00 Median :0.9968
## Mean :0.08747 Mean :15.87 Mean : 46.47 Mean :0.9967
## 3rd Qu.:0.09000 3rd Qu.:21.00 3rd Qu.: 62.00 3rd Qu.:0.9978
## Max. :0.61100 Max. :72.00 Max. :289.00 Max. :1.0037
## pH sulphates alcohol quality
## Min. :2.740 Min. :0.3300 Min. : 8.40 Min. :3.000
## 1st Qu.:3.210 1st Qu.:0.5500 1st Qu.: 9.50 1st Qu.:5.000
## Median :3.310 Median :0.6200 Median :10.20 Median :6.000
## Mean :3.311 Mean :0.6581 Mean :10.42 Mean :5.636
## 3rd Qu.:3.400 3rd Qu.:0.7300 3rd Qu.:11.10 3rd Qu.:6.000
## Max. :4.010 Max. :2.0000 Max. :14.90 Max. :8.000
```

```
model <- lm(pH ~ . -citric_acid, data = redwine)
```

```
summary(model)
```

```
##
## Call:
```

```
## lm(formula = pH ~ . - citric_acid, data = redwine)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.33633 -0.05101 -0.00120  0.05177  0.46071
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -6.057e+01  2.321e+00 -26.094 < 2e-16 ***
## fixed_acidity  -9.859e-02  2.033e-03 -48.506 < 2e-16 ***
## volatile_acidity  2.142e-02  1.377e-02  1.555  0.1201
## residual_sugar  -2.576e-02  1.852e-03 -13.912 < 2e-16 ***
## chlorides      -5.385e-01  5.160e-02 -10.436 < 2e-16 ***
## free_sulfur_dioxide  1.654e-03  2.777e-04  5.954 3.21e-09 ***
## total_sulfur_dioxide -7.951e-04  9.050e-05 -8.785 < 2e-16 ***
## density        6.435e+01  2.326e+00 27.663 < 2e-16 ***
## sulphates      -7.082e-02  1.515e-02 -4.675 3.19e-06 ***
## alcohol         7.294e-02  3.031e-03 24.066 < 2e-16 ***
## quality        -6.942e-03  3.279e-03 -2.117  0.0344 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0848 on 1588 degrees of freedom
## Multiple R-squared:  0.7002, Adjusted R-squared:  0.6983
## F-statistic: 370.8 on 10 and 1588 DF, p-value: < 2.2e-16
```

```
library(tidyverse) # For data manipulation
library(broom)
model_results <- tidy(model)
model_results %>% select(term, estimate, std.error, statistic, p.value)
```

```
## # A tibble: 11 x 5
##   term                estimate std.error statistic  p.value
##   <chr>              <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)       -60.6        2.32     -26.1 3.33e-125
## 2 fixed_acidity     -0.0986     0.00203   -48.5 9.89e-316
## 3 volatile_acidity   0.0214     0.0138     1.56 1.20e- 1
## 4 residual_sugar    -0.0258     0.00185   -13.9 1.33e- 41
## 5 chlorides         -0.538      0.0516    -10.4 1.05e- 24
## 6 free_sulfur_dioxide 0.00165    0.000278    5.95 3.21e- 9
## 7 total_sulfur_dioxide -0.000795  0.0000905   -8.78 3.97e- 18
## 8 density           64.3        2.33     27.7 8.29e-138
## 9 sulphates         -0.0708     0.0152    -4.67 3.19e- 6
## 10 alcohol           0.0729     0.00303    24.1 2.31e-109
## 11 quality          -0.00694    0.00328    -2.12 3.44e- 2
```

#Explanation:

- The summary(model) command provides the regression results, showing the coefficient estimates, their standard errors, t-values, and p-values for each predictor.

```
significant_predictors <- model_results %>% filter(p.value < 0.05)
significant_predictors
```

```
## # A tibble: 10 x 5
##   term                estimate std.error statistic  p.value
```

##	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
## 1	(Intercept)	-60.6	2.32	-26.1	3.33e-125
## 2	fixed_acidity	-0.0986	0.00203	-48.5	9.89e-316
## 3	residual_sugar	-0.0258	0.00185	-13.9	1.33e- 41
## 4	chlorides	-0.538	0.0516	-10.4	1.05e- 24
## 5	free_sulfur_dioxide	0.00165	0.000278	5.95	3.21e- 9
## 6	total_sulfur_dioxide	-0.000795	0.0000905	-8.78	3.97e- 18
## 7	density	64.3	2.33	27.7	8.29e-138
## 8	sulphates	-0.0708	0.0152	-4.67	3.19e- 6
## 9	alcohol	0.0729	0.00303	24.1	2.31e-109
## 10	quality	-0.00694	0.00328	-2.12	3.44e- 2

(1) The significant_predictors will show only those predictors that have a p-value less than 0.05, which suggests that these predictors have a statistically significant relationship with the response (pH). * All predictors except volatile_acidity, are significant.

```
coef <- model_results %>% filter(term == "free_sulfur_dioxide") %>% pull(estimate)
coef
```

```
## [1] 0.001653707
```

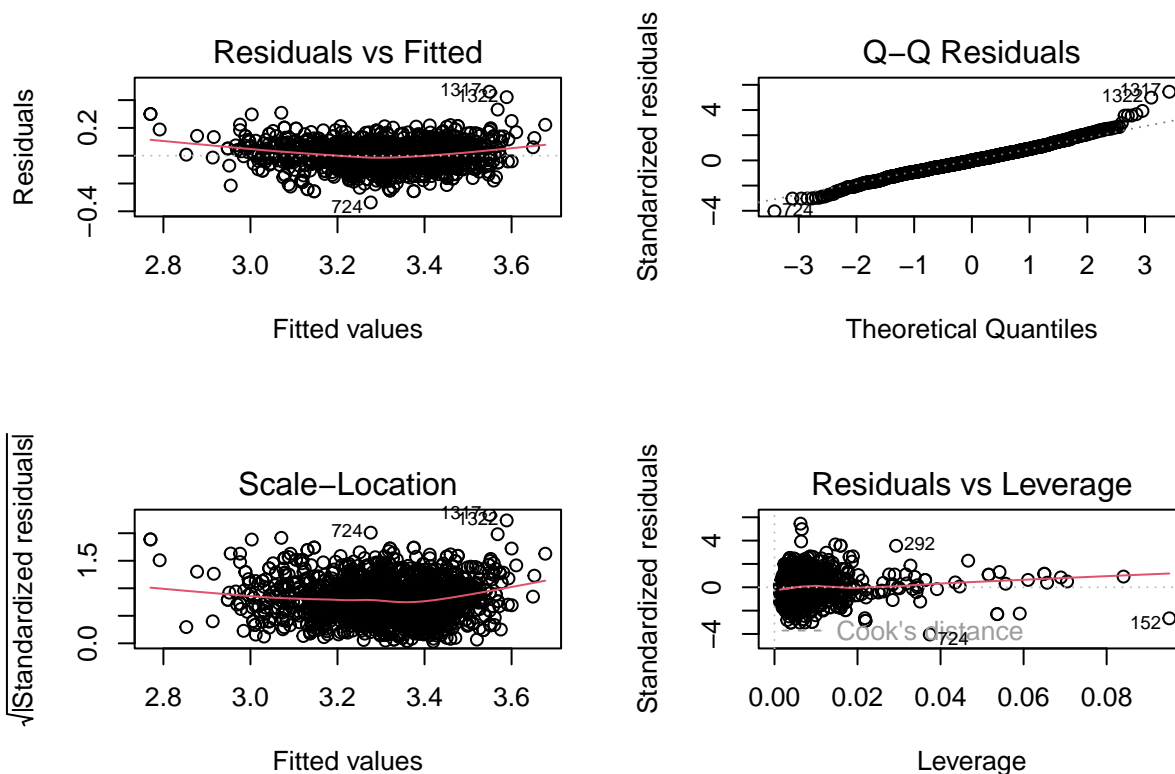
(2) The coefficient for free_sulfur_dioxide is extracted and stored in coef= 0.001653707. It represents the expected change in pH for each unit increase in free_sulfur_dioxide, holding all other variables constant

b.

(6 points) Produce diagnostic plots of the linear regression fit. Comment on any problems you see with the fit. Do the residual plots suggest any unusually large outliers? Does the leverage plot identify any observations with unusually high leverage?

#Solution:

```
par(mfrow = c(2, 2))
plot(model)
```



```
par(mfrow = c(1, 1))
```

#Explanation:

Residuals vs. Fitted: Random scatter indicates good fit; patterns suggest non-linearity in the model.

Q-Q: Residuals should follow a straight line to confirm normal distribution.

Scale-Location: A flat horizontal line shows equal variance; a funnel shape suggests heteroscedasticity.

Residuals vs. Leverage: Points beyond Cook's distance indicate influential data affecting the model.

- c. (6 points) Fit at least 3 linear regression models (exploring interaction effects) with alcohol as the response and some combination of other variables as predictors. Do any interactions appear to be statistically significant?

```
library(dplyr)
```

```
model1 <- lm(alcohol ~ sulphates * density, data = redwine)
summary(model1)
```

```
##
## Call:
## lm(formula = alcohol ~ sulphates * density, data = redwine)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1334 -0.6790 -0.1242  0.5633  4.5359
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      266.26      53.96   4.934 8.89e-07 ***
## sulphates         56.97      79.44   0.717  0.473
```

```

## density          -257.39      54.13  -4.755 2.17e-06 ***
## sulphates:density  -56.06      79.68  -0.704  0.482
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9082 on 1595 degrees of freedom
## Multiple R-squared:  0.275, Adjusted R-squared:  0.2737
## F-statistic: 201.7 on 3 and 1595 DF, p-value: < 2.2e-16
model2 <- lm(alcohol ~ fixed_acidity * volatile_acidity, data = redwine)
summary(model2)

##
## Call:
## lm(formula = alcohol ~ fixed_acidity * volatile_acidity, data = redwine)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1243 -0.8375 -0.1888  0.6659  4.6257
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      10.94390    0.41331   26.479 <2e-16 ***
## fixed_acidity       0.02525    0.04778    0.529  0.5972
## volatile_acidity    0.29541    0.77954    0.379  0.7048
## fixed_acidity:volatile_acidity -0.20571    0.09343   -2.202  0.0278 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.036 on 1595 degrees of freedom
## Multiple R-squared:  0.05757, Adjusted R-squared:  0.05579
## F-statistic: 32.48 on 3 and 1595 DF, p-value: < 2.2e-16
model3 <- lm(alcohol ~ chlorides * free_sulfur_dioxide, data = redwine)
summary(model3)

##
## Call:
## lm(formula = alcohol ~ chlorides * free_sulfur_dioxide, data = redwine)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1357 -0.8719 -0.2197  0.6476  4.5641
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      10.811207    0.102920  105.045 < 2e-16 ***
## chlorides        -3.262980    1.012258   -3.223  0.00129 **
## free_sulfur_dioxide  0.002835    0.005390    0.526  0.59893
## chlorides:free_sulfur_dioxide -0.106259    0.051995   -2.044  0.04115 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.036 on 1595 degrees of freedom
## Multiple R-squared:  0.05602, Adjusted R-squared:  0.05425

```

```
## F-statistic: 31.55 on 3 and 1595 DF, p-value: < 2.2e-16
```

```
#Explanation:
```

Model 1: Interaction between sulphates and density. Model 2: Interaction between fixed_acidity and volatile_acidity. Model 3: Interaction between chlorides and free_sulfur_dioxide.

P (model 1) = 0.482 P (model 2) = 0.0278 P (model 3) = 0.04115

model 2 and model 3 are statistically significant with p values less than 0.05. Where as, model 1 is not significant.

```
#Question 2:
```

(30 points) This problem involves the Boston data set, which can be loaded from library MASS in R and is also made available in the Datasets for Assignments module on Canvas (boston.csv). We will now try to predict per capita crime rate (crim) using the other variables in this data set. In other words, per capita crime rate is the response, and the other variables are the predictors.

```
Boston=read.csv("/Users/sharathkarnati/Desktop/DS/Assignment 5/boston.csv", sep=" ", header = TRUE )
Boston <- na.omit(Boston)
summary(Boston)
```

```
## crim.zn.indus.chas.nox.rm.age.dis.rad.tax.ptratio.black.lstat.medv
## Length:506
## Class :character
## Mode :character
```

- a. (6 points) For each predictor, fit a simple linear regression model to predict the response. Include the code, but not the output for all the models in your solution.

```
#Solution:
```

```
library(MASS)
```

```
data(Boston)
```

```
#linear regression
```

```
#zn
```

```
lm_zn <- lm(crim ~ zn , data = Boston)
```

```
#indus
```

```
lm_indus <- lm(crim ~ indus , data = Boston)
```

```
#chas
```

```
lm_chas <- lm(crim ~ chas , data = Boston)
```

```
#nox
```

```
lm_nox <- lm(crim ~ nox , data = Boston)
```

```
#rm
```

```
lm_rm <- lm(crim ~ rm , data = Boston)
```

```
#age
```

```
lm_age <- lm(crim ~ age , data = Boston)
```

```
#dis
```

```
lm_dis <- lm(crim ~ dis , data = Boston)
```

```

#rad
lm_rad <- lm(crim ~ rad , data = Boston)

#tax
lm_tax <- lm(crim ~ tax , data = Boston)

#ptratio
lm_ptratio <- lm(crim ~ ptratio , data = Boston)

#black
lm_black <- lm(crim ~ black , data = Boston)

#lstat
lm_lstat <- lm(crim ~ lstat , data = Boston)

#medv
lm_medv <- lm(crim ~ medv , data = Boston)

```

- b. (6 points) In which of the models is there a statistically significant association between the predictor and the response? Considering the meaning of each variable, discuss the relationship between crim and each of the predictors nox, chas, rm, dis and medv. How do these relationships differ?

#Solution :

There is a statistically significant link between the predictor and the response for all models, except for “chas.”

Here’s what each variable means:

crim: the crime rate per person in each town. nox: the concentration of nitrogen oxides in the air. chas: a dummy variable that indicates whether the area is near the Charles River (1 if yes, 0 if no). rm: the average number of rooms per home. dis: the average distance to five major job centers in Boston. medv: the median value of homes, in thousands of dollars. The other variables all relate to different aspects of Boston—environment, housing, or employment statistics—that tend to show some relationship with crime rates, except for “chas,” which doesn’t have a strong connection with the other variables.

- c. (6 points) Fit a multiple regression model to predict the response using all the predictors. Describe your results. For which predictors can we reject the null hypothesis $H_0 : \beta_j = 0$?

Solution :

```

lm_all <- lm(crim ~. , data = Boston )
summary(lm_all)

##
## Call:
## lm(formula = crim ~ ., data = Boston)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.924 -2.120 -0.353  1.019 75.051
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  17.033228   7.234903   2.354 0.018949 *

```

```
## zn          0.044855    0.018734    2.394 0.017025 *
## indus      -0.063855    0.083407   -0.766 0.444294
## chas       -0.749134    1.180147   -0.635 0.525867
## nox        -10.313535    5.275536   -1.955 0.051152 .
## rm         0.430131    0.612830    0.702 0.483089
## age        0.001452    0.017925    0.081 0.935488
## dis        -0.987176    0.281817   -3.503 0.000502 ***
## rad        0.588209    0.088049    6.680 6.46e-11 ***
## tax        -0.003780    0.005156   -0.733 0.463793
## ptratio    -0.271081    0.186450   -1.454 0.146611
## black      -0.007538    0.003673   -2.052 0.040702 *
## lstat      0.126211    0.075725    1.667 0.096208 .
## medv       -0.198887    0.060516   -3.287 0.001087 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.439 on 492 degrees of freedom
## Multiple R-squared:  0.454, Adjusted R-squared:  0.4396
## F-statistic: 31.47 on 13 and 492 DF,  p-value: < 2.2e-16
```

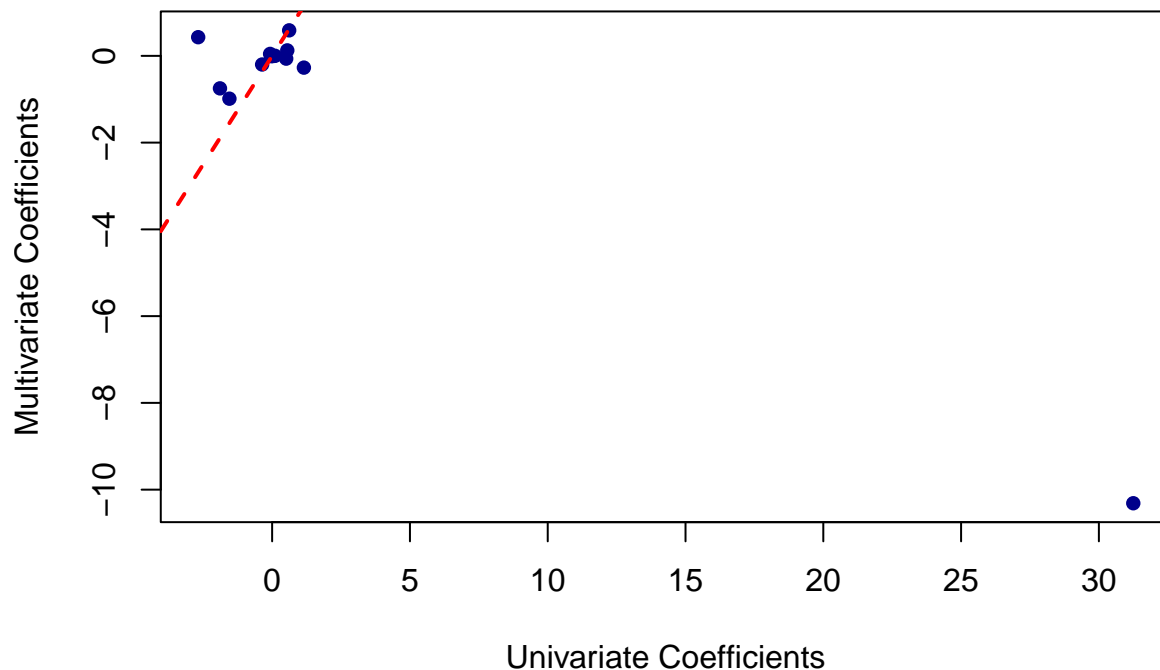
We have achieved the above results to predict the response by using all the predictors and the null hypothesis can be neglected for the predictors “zn”, “dis”, “rad”, “black”, and medv as their p value is less than 0.05

- d. (6 points) How do your results from (a) compare to your results from (c)? You can present this comparison as a plot or as a table or any other form of comparison you deem fit.

```
univariate_pred <- c(
  lm_zn$coefficients[2],
  lm_indus$coefficients[2],
  lm_chas$coefficients[2],
  lm_nox$coefficients[2],
  lm_rm$coefficients[2],
  lm_age$coefficients[2],
  lm_dis$coefficients[2],
  lm_rad$coefficients[2],
  lm_tax$coefficients[2],
  lm_ptratio$coefficients[2],
  lm_black$coefficients[2],
  lm_lstat$coefficients[2],
  lm_medv$coefficients[2]
)
multivariate_pred <- lm_all$coefficients[2:14]

plot(
  univariate_pred, multivariate_pred,
  xlab = "Univariate Coefficients",
  ylab = "Multivariate Coefficients",
  main = "Univariate vs. Multivariate Coefficients",
  pch = 16,    # Filled circle points
  col = "darkblue" # Set color of points
)
abline(a = 0, b = 1, col = "red", lty = 2, lwd = 2)
```


Univariate vs. Multivariate Coefficients



We can observe from the graph that the values for the coefficient that depict the variable is different when modeled alone compared to model having all together.

e.(6 points) Is there evidence of non-linear association between the predictors age and tax and the response crim? To answer this question, for each predictor (age and tax), fit a model of the form: $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3$ Hint: use the `poly()` function in R. Use the model to assess the extent of non-linear association.

#Solution :

```
poly_fit_age <- lm(formula = crim ~ poly(age, 3), data = Boston)
anova(lm_age, poly_fit_age)
```

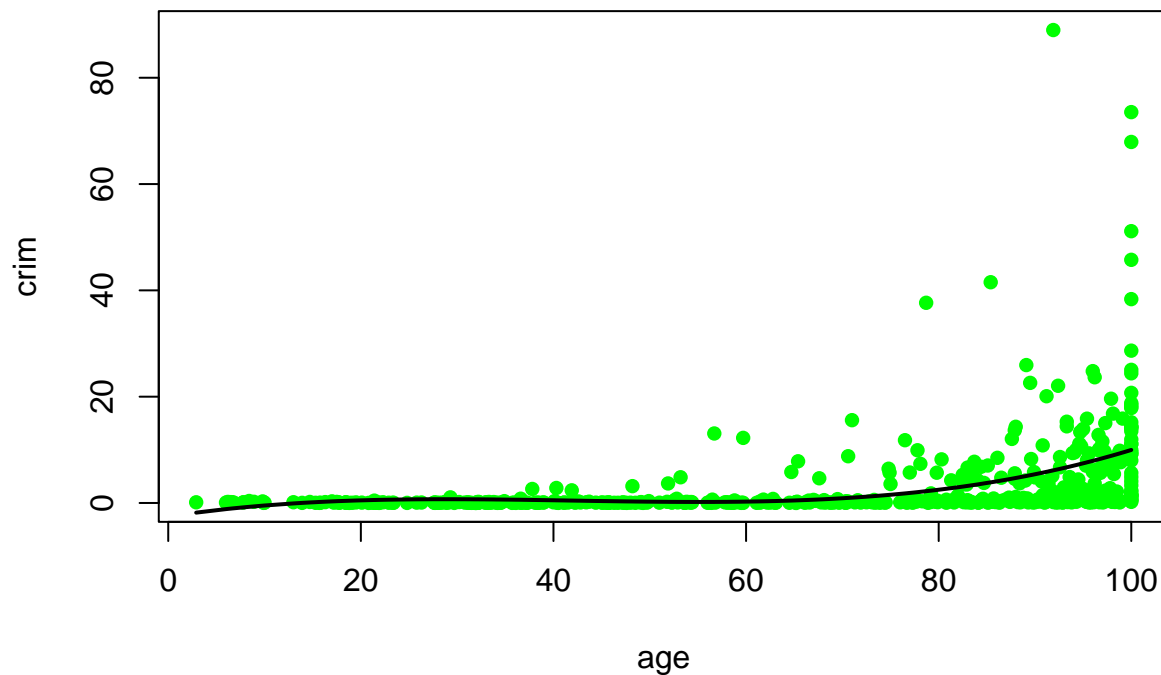
```
## Analysis of Variance Table
##
## Model 1: crim ~ age
## Model 2: crim ~ poly(age, 3)
##   Res.Df  RSS Df Sum of Sq    F    Pr(>F)
## 1     504 32714
## 2     502 30853   2     1861 15.14 4.125e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
poly_fit_tax <- lm(formula = crim ~ poly(tax, 3), data = Boston)
anova(lm_tax, poly_fit_tax)
```

```
## Analysis of Variance Table
##
## Model 1: crim ~ tax
## Model 2: crim ~ poly(tax, 3)
##   Res.Df  RSS Df Sum of Sq    F    Pr(>F)
## 1     504 24674
## 2     502 23581   2     1093.5 11.64 1.144e-05 ***
```

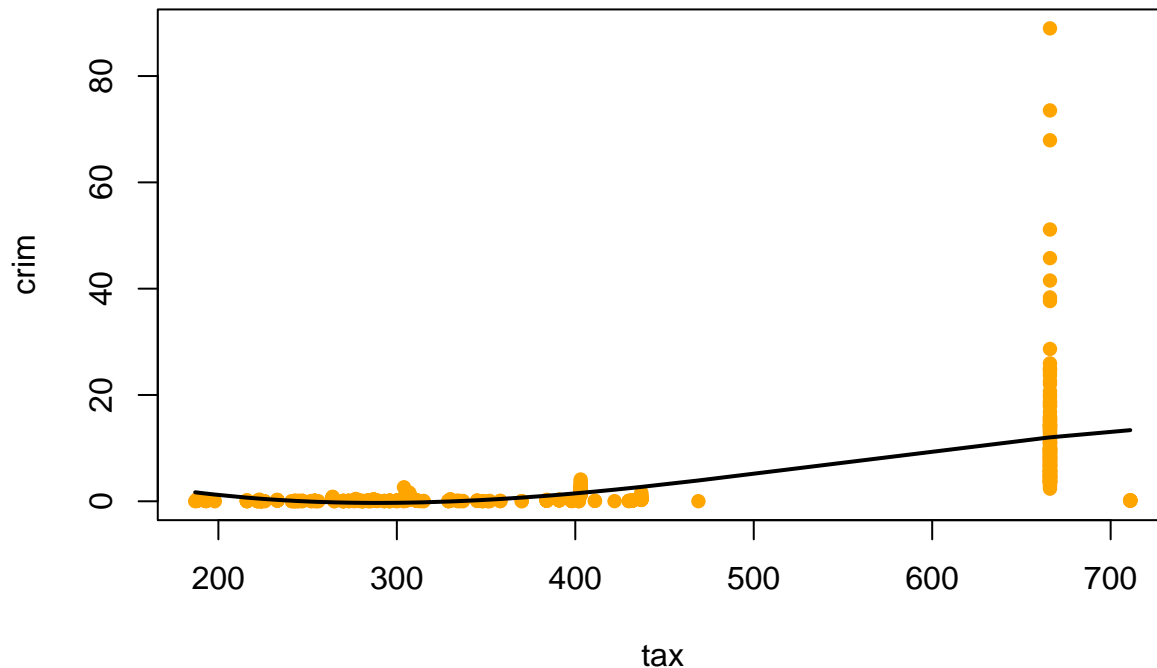
```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
plot(Boston$age, Boston$crim, main = "Polynomial Fit for age vs crim", xlab = "age",
     ylab = "crim", col = "green", pch = 16)
lines(sort(Boston$age), fitted(poly_fit_age)[order(Boston$age)],
      col = "black", lwd = 2)
```

Polynomial Fit for age vs crim



```
plot(Boston$tax, Boston$crim, main = "Polynomial Fit for tax vs crim", xlab = "tax",
     ylab = "crim", col = "orange", pch = 16)
lines(sort(Boston$tax), fitted(poly_fit_tax)[order(Boston$tax)],
      col = "black", lwd = 2)
```

Polynomial Fit for tax vs crim



The code fits a cubic polynomial for both age and tax to see if the non-linear relationship with crim (crime rate) improves the model. An ANOVA test is used to compare the linear and polynomial models, checking if the polynomial terms add any significant value. Scatter plots show the relationship between age, tax, and crim, with a black line representing the polynomial fit to highlight any non-linear patterns in the data.

#Question 3:

- 3) (12 points) Suppose we collect data for a group of students in a statistics class with variables: X_1 = hours studied, X_2 = undergrad GPA, X_3 = PSQI score (a sleep quality index), and Y = receive an A. We fit a logistic regression and produce estimated coefficient, $\beta_0 = -8$, $\beta_1 = 0.1$, $\beta_2 = 1$, $\beta_3 = -.04$.
 - a. (4 points) Estimate the probability that a student who studies for 32 h, has a PSQI score of 11 and has an undergrad GPA of 3.0 gets an A in the class. Show your work.

#Solution:

```
#Given ;
beta_0 <- -8
beta_1 <- 0.1
beta_2 <- 1
beta_3 <- -0.04
hours_studied <- 32
gpa <- 3.0
psqi <- 11

#Solving the problem :

log_odds <- beta_0 + beta_1 * hours_studied + beta_2 * gpa + beta_3 * psqi

probability <- exp(log_odds) / (1 + exp(log_odds))
```

```
probability
```

```
## [1] 0.09621554
```

Probability of getting A for the student is 0.09621554

- b. (4 points) How many hours would the student in part (a) need to study to have a 65 % chance of getting an A in the class? Show your work.

#Solution:

Logistic Regression Equation

The logistic regression equation is:

$$P = \frac{e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3)}}{1 + e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3)}}$$

To isolate the number of hours studied X_1 , we rearrange the equation as follows:

$$\log\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

Now, solving for X_1 :

$$X_1 = \frac{\log\left(\frac{P}{1-P}\right) - \beta_0 - \beta_2 X_2 - \beta_3 X_3}{\beta_1}$$

Where: - $P = 0.65$ (probability of getting an A), - $\beta_0 = -8$, - $\beta_1 = 0.1$ (hours studied), - $\beta_2 = 1$ (undergrad GPA), - $\beta_3 = -0.04$ (PSQI score), - $X_2 = 3.0$ (GPA), - $X_3 = 11$ (PSQI score).

R Code

```
# Given
beta_0 <- -8
beta_1 <- 0.1
beta_2 <- 1
beta_3 <- -0.04
gpa <- 3.0
psqi <- 11
P <- 0.65

#Solving the problem :

log_odds <- log(P / (1 - P))

hours_studied_needed <- (log_odds - beta_0 - beta_2 * gpa - beta_3 * psqi) / beta_1

hours_studied_needed
```

```
## [1] 60.59039
```

Therefore, he needs to study for 60.59039 hrs to get a A or 65 percent above.

- c. (4 points) How many hours would a student with a 3.0 GPA and a PSQI score of 3 need to study to have a 60 % chance of getting an A in the class? Show your work.

#Solution :

Logistic Regression Equation

The logistic regression equation is given by:

$$P = \frac{e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3)}}{1 + e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3)}}$$

To isolate the number of hours studied X_1 , we rearrange the equation as follows:

$$\log\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

Now, solving for X_1 :

$$X_1 = \frac{\log\left(\frac{P}{1-P}\right) - \beta_0 - \beta_2 X_2 - \beta_3 X_3}{\beta_1}$$

Where: - $P = 0.60$ (probability of getting an A), - $\beta_0 = -8$, - $\beta_1 = 0.1$ (hours studied), - $\beta_2 = 1$ (undergrad GPA), - $\beta_3 = -0.04$ (PSQI score), - $X_2 = 3.0$ (GPA), - $X_3 = 3$ (PSQI score).

R Code :

```
# Given
beta_0 <- -8
beta_1 <- 0.1
beta_2 <- 1
beta_3 <- -0.04
gpa <- 3.0
psqi <- 3
P <- 0.60

#Solving:

log_odds <- log(P / (1 - P))

hours_studied_needed <- (log_odds - beta_0 - beta_2 * gpa - beta_3 * psqi) / beta_1

hours_studied_needed

## [1] 55.25465
```

the caluclated number of hours required to study for getting 3.0 gpa is 55.25465 hrs.