Assignment 5

Athul Jose P
11867566

School of Electrical Engineering and Computer Science

Washington State University

CptS 575 Data Science

1.a.

```
# loading dataset
df <- read.csv('winequality-red.csv', header = TRUE)

# linear regression with pH except citric acid
model <- lm(pH ~ . - citric_acid, data = df)

# summary
model_summary <- summary(model)

#print summary
model_summary
```

Call:
lm(formula = pH ~ . - citric_acid, data = df)

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

1.a.i

To identify the predictors which are statistically significant, p-value is taken as a measure. If p-value is less than 0.05, the predictor is considered statistically significant, Otherwise it is not statistically not significant.

Significant Predictors ($p < 0.05$):

```
# Extract the coefficients and p-values
coeff_df <- as.data.frame(model_summary$coefficients)

# Filter predictors with p-values < 0.05
significant_pred <- coeff_df[coeff_df$`Pr(>|t|)` < 0.05, ]
print(significant_pred)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-6.056702e+01	2.321123e+00	-26.093842	3.325214e-125
fixed_acidity	-9.859226e-02	2.032595e-03	-48.505618	9.885331e-316
residual_sugar	-2.575873e-02	1.851521e-03	-13.912196	1.327582e-41
chlorides	-5.384863e-01	5.159810e-02	-10.436166	1.045480e-24
free_sulfur_dioxide	1.653707e-03	2.777318e-04	5.954333	3.208399e-09
total_sulfur_dioxide	-7.950509e-04	9.050109e-05	-8.784988	3.971243e-18
density	6.434507e+01	2.326051e+00	27.662789	8.291055e-138
sulphates	-7.082418e-02	1.515092e-02	-4.674578	3.194769e-06
alcohol	7.294214e-02	3.030915e-03	24.066045	2.305650e-109
quality	-6.942504e-03	3.278813e-03	-2.117383	3.438240e-02

Non-significant Predictors ($p >= 0.05$):

```
# Filter predictors with p-values >= 0.05
non_significant_pred <- coeff_df[coeff_df$`Pr(>|t|)` >= 0.05, ]
print(non_significant_pred)
```

	Estimate	Std. Error	t value	Pr(>|t|)
volatile_acidity	0.02141973	0.01377346	1.555145	0.1201107

1.a.ii

The coefficient for the free_sulfur_dioxide variable in regression model is 1.653707e-3. In simple terms, for every 1 unit increase in free sulfur dioxide, the predicted pH increases by 0.001653 units, keeping all other variables constant.

1.b.

```
# plot diagnostic plots
par(mfrow = c(2, 2))
plot(model)
```
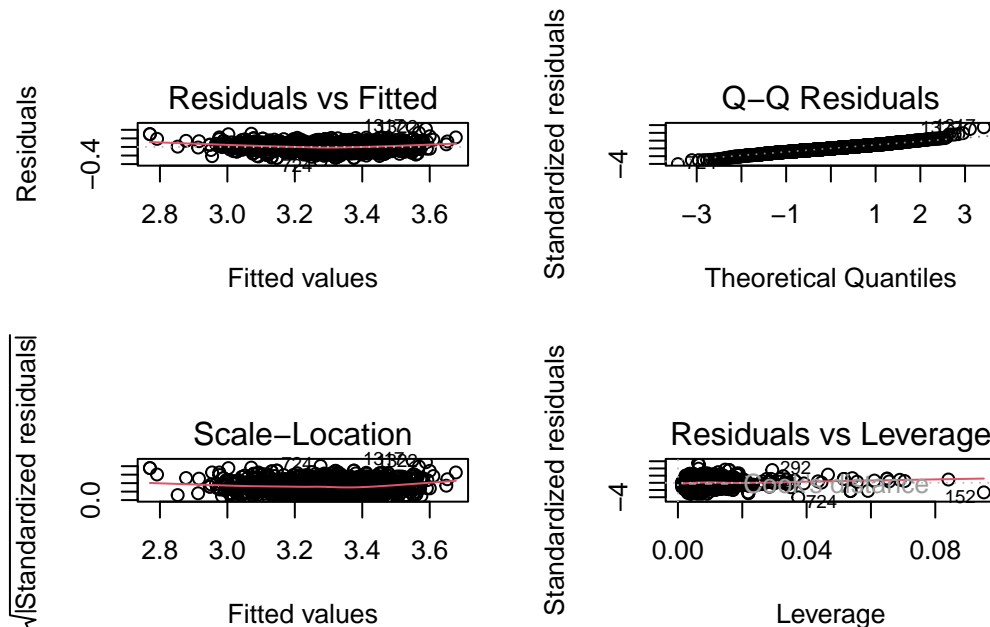

Residuals vs. Fitted: The residuals are mostly scattered evenly around the horizontal line at zero, suggesting that the linear relationship between the predictors and the response is appropriate. However, there are a few points that stand out slightly (e.g., observations 1313 and 152), which may indicate potential outliers.

Q-Q Residual: Most points lie on the 45-degree line, which suggests that the residuals follow a normal distribution fairly well. However, the points at the far tails (e.g., 1313) deviate slightly from the line, indicating the presence of some outliers.

Scale-Location: The points are fairly evenly spread along the horizontal line, which suggests that the variance of residuals is relatively constant. A few points (e.g., 1313) stand out, which might indicate mild issues with variance at certain fitted values.

Residuals vs Leverage: Most points have low leverage, clustered to the left near zero. Observations 152 and 292 as high leverage points

1.c.

```
# Model 1: Interaction between sulphates and residual sugar
model1 <- lm(alcohol ~ sulphates * residual_sugar, data = df)
summary(model1)
```

```
Call:
lm(formula = alcohol ~ sulphates * residual_sugar, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-2.1013 -0.8564 -0.2515  0.6433  4.1519

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               10.06280    0.24466  41.130   <2e-16 ***
sulphates                  0.42589    0.36289   1.174    0.241
residual_sugar            -0.01290    0.09200  -0.140    0.889
sulphates:residual_sugar   0.06735    0.13688   0.492    0.623
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.061 on 1595 degrees of freedom
Multiple R-squared:  0.01064,    Adjusted R-squared:  0.008776
F-statistic: 5.716 on 3 and 1595 DF,  p-value: 0.0006861
```

No significant interactions are found between sulphates and residual_sugar in this model. Additionally, the individual predictors themselves (sulphates and residual_sugar) are not statistically significant either.

```
# Model 2: Interaction between density and total sulfur dioxide
model2 <- lm(alcohol ~ density * total_sulfur_dioxide, data = df)
summary(model2)
```

```
Call:
lm(formula = alcohol ~ density * total_sulfur_dioxide, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-1.8573 -0.6932 -0.1104  0.5500  4.8527

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                   264.4495    20.4759  12.915   <2e-16 ***
density                      -254.5926    20.5474 -12.391   <2e-16 ***
total_sulfur_dioxide            0.3638     0.3288   1.106    0.269
density:total_sulfur_dioxide   -0.3706     0.3300  -1.123    0.262
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9077 on 1595 degrees of freedom
Multiple R-squared:  0.2759,    Adjusted R-squared:  0.2746
F-statistic: 202.6 on 3 and 1595 DF,  p-value: < 2.2e-16
```

The interaction effect between density and total_sulfur_dioxide is not statistically significant (p-value = 0.262). density has a strong, statistically significant effect on alcohol, but total_sulfur_dioxide does not.

```
# Model 3: Interaction between volatile acidity and chlorides
model3 <- lm(alcohol ~ volatile_acidity * chlorides, data = df)
summary(model3)
```

```
Call:
lm(formula = alcohol ~ volatile_acidity * chlorides, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-2.3288 -0.8104 -0.1876  0.6497  4.3296

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                 11.4597     0.1761  65.088  < 2e-16 ***
volatile_acidity            -1.1789     0.3245  -3.632  0.00029 ***
chlorides                   -5.0576     1.8583  -2.722  0.00657 **
volatile_acidity:chlorides   0.5972     3.3869   0.176  0.86006
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.021 on 1595 degrees of freedom
Multiple R-squared:  0.08468,   Adjusted R-squared:  0.08295
F-statistic: 49.18 on 3 and 1595 DF,  p-value: < 2.2e-16
```

The interaction effect between volatile_acidity and chlorides is not statistically significant (p-value = 0.86006). However, both volatile_acidity and chlorides individually have statistically significant effects on alcohol content.

2.a.

```
# loading MASS library and Boston dataset
library(MASS)
```

Attaching package: 'MASS'

The following object is masked from 'package:dplyr':

    select

```
data("Boston")

# predictors
predictors <- setdiff(names(Boston), "crim")
predictors
```

```
 [1] "zn"      "indus"   "chas"    "nox"     "rm"      "age"      "dis"
 [8] "rad"     "tax"     "ptratio" "black"   "lstat"   "medv"
```

```
# fit a simple linear regression model
model_zn      <- lm(crim ~ zn, data = Boston)
model_indus   <- lm(crim ~ indus, data = Boston)
model_chas    <- lm(crim ~ chas, data = Boston)
model_nox     <- lm(crim ~ nox, data = Boston)
model_rm      <- lm(crim ~ rm, data = Boston)
model_age     <- lm(crim ~ age, data = Boston)
model_dis     <- lm(crim ~ dis, data = Boston)
model_rad     <- lm(crim ~ rad, data = Boston)
model_tax     <- lm(crim ~ tax, data = Boston)
model_ptratio <- lm(crim ~ ptratio, data = Boston)
model_black   <- lm(crim ~ black, data = Boston)
model_lstat   <- lm(crim ~ lstat, data = Boston)
model_medv    <- lm(crim ~ medv, data = Boston)
```

2.b.

```
Call:
lm(formula = crim ~ nox, data = Boston)

Residuals:
    Min      1Q  Median      3Q     Max
-12.371  -2.738  -0.974   0.559  81.728

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -13.720      1.699  -8.073 5.08e-15 ***
nox           31.249      2.999  10.419  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.81 on 504 degrees of freedom
Multiple R-squared:  0.1772,    Adjusted R-squared:  0.1756
F-statistic: 108.6 on 1 and 504 DF,  p-value: < 2.2e-16
```

A positive and significant relationship suggests that areas with higher pollution levels (nox) tend to have higher crime rates. This makes sense since higher nox concentrations are often found in urban and industrial areas, which may also have higher crime rates compared to suburban or rural areas.

```
Call:
lm(formula = crim ~ chas, data = Boston)

Residuals:
   Min     1Q Median     3Q    Max
-3.738 -3.661 -3.435  0.018 85.232

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.7444     0.3961   9.453   <2e-16 ***
chas         -1.8928     1.5061  -1.257    0.209
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 8.597 on 504 degrees of freedom
Multiple R-squared:  0.003124,  Adjusted R-squared:  0.001146
F-statistic: 1.579 on 1 and 504 DF,  p-value: 0.2094
```

The negative coefficient suggests a potential trend where areas near the Charles River might have lower crime rates. However, this effect is not statistically significant. Even though living near the river might indicate more desirable neighborhoods, in this analysis, the effect of chas on crim is not strong enough to be considered statistically significant.

summary(model_rm)

```
Call:
lm(formula = crim ~ rm, data = Boston)

Residuals:
   Min     1Q Median     3Q    Max
-6.604 -3.952 -2.654  0.989 87.197

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   20.482      3.365   6.088 2.27e-09 ***
rm            -2.684      0.532  -5.045 6.35e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.401 on 504 degrees of freedom
Multiple R-squared:  0.04807,   Adjusted R-squared:  0.04618
F-statistic: 25.45 on 1 and 504 DF,  p-value: 6.347e-07
```

There is a significant negative relationship between the average number of rooms (rm) and per capita crime rate (crim). This suggests that areas with larger homes tend to have lower crime rates, likely reflecting the fact that wealthier or suburban areas generally experience less crime.

summary(model_dis)

```
Call:
lm(formula = crim ~ dis, data = Boston)

Residuals:
```

```
    Min      1Q Median      3Q     Max
-6.708 -4.134 -1.527   1.516 81.674


Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   9.4993     0.7304  13.006   <2e-16 ***
dis          -1.5509     0.1683  -9.213   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 7.965 on 504 degrees of freedom
Multiple R-squared:  0.1441,    Adjusted R-squared:  0.1425
F-statistic: 84.89 on 1 and 504 DF,  p-value: < 2.2e-16
```

The model shows a statistically significant negative relationship between dis and crim. As distance from employment centers increases, crime rates tend to decrease, suggesting that suburban or rural areas have lower crime rates compared to urban areas.

```
summary(model_medv)
```

```
Call:
lm(formula = crim ~ medv, data = Boston)

Residuals:
    Min      1Q Median      3Q     Max
-9.071 -4.022 -2.343   1.298 80.957

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 11.79654    0.93419   12.63   <2e-16 ***
medv        -0.36316    0.03839   -9.46   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.934 on 504 degrees of freedom
Multiple R-squared:  0.1508,    Adjusted R-squared:  0.1491
F-statistic: 89.49 on 1 and 504 DF,  p-value: < 2.2e-16
```

There is a statistically significant negative relationship between the median value of homes (medv) and the crime rate (crim). As the median value of homes increases, the crime rate decreases, suggesting that wealthier neighborhoods tend to have lower crime rates.

Generally, the analysis highlights that crime tends to increse with urbanization (pollution) and decrease with affluence and suburbanization (more rooms, higher home values, greater distance from city centers)

2.c.

```
#  multiple regression model with all predictors
model_full <- lm(crim ~ ., data = Boston)

# print summary
summary(model_full)
```

```
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
```

The multiple linear regression model for predicting the per capita crime rate (crim) reveals several significant predictors. Highway accessibility (rad) shows a strong positive relationship with crime, indicating that areas with greater highway access tend to have higher crime rates. Conversely, distance to employment centers (dis) and zoned residential land (zn) exhibit significant negative effects, suggesting that suburban or rural areas experience lower crime. Additionally, higher property values (medv) and greater proportion of Black residents (black) are associated with lower crime rates. These findings align with general expectations that affluent, suburban areas with fewer urban stressors experience less crime.

However, some variables, such as pollution levels (nox) and percentage of lower-status individuals (lstat), show marginal significance, suggesting complex or weak relationships with crime. Predictors like proximity to the Charles River (chas), average number of rooms (rm), tax rates, and pupil-teacher ratio (ptratio) are not statistically significant, indicating limited direct influence on crime rates. The model explains 45.4% of the variance in the crime rate, providing a moderate fit, with the overall model being statistically significant. These results highlight how a combination of urbanization, infrastructure, property values, and social factors contribute to crime patterns.

For these five predictors zn, dis, rad, black, and medv, null hypothesis can be rejected meaning they have a statistically significant effect on the crime rate. All other predictors in the model have p-values 0.05, so we fail to reject the null hypothesis for those predictors.

2.d.

```
library(ggplot2)

# exclude intercept
multiple_pvalues <- summary(model_full)$coefficients[-1, "Pr(>|t|)"]

# get p values from simple
get_simple_pvalue <- function(predictor) {
  model <- lm(as.formula(paste("crim ~", predictor)), data = Boston)
  summary(model)$coefficients[2, "Pr(>|t|)"]
}
predictors <- names(Boston)[names(Boston) != "crim"]
simple_pvalues <- sapply(predictors, get_simple_pvalue)

# collect data
comparison_df <- data.frame(
  Predictor = predictors,
  Simple_Regression = simple_pvalues,
  Multiple_Regression = multiple_pvalues
)
```

```
# df for plotting
comparison_df_long <- data.frame(
  Predictor = rep(comparison_df$Predictor, 2),
  Model = c(rep("Simple Regression", length(predictors)),
            rep("Multiple Regression", length(predictors))),
  p_value = c(comparison_df$Simple_Regression, comparison_df$Multiple_Regression)
)

# plotting bar plot
ggplot(comparison_df_long, aes(x = Predictor, y = p_value, fill = Model)) +
  geom_bar(stat = "identity", position = "dodge") +
  scale_y_log10() +
  labs(title = "Comparison of P-values: Simple vs. Multiple Regression",
       y = "P-value (log scale)", x = "Predictor")
```
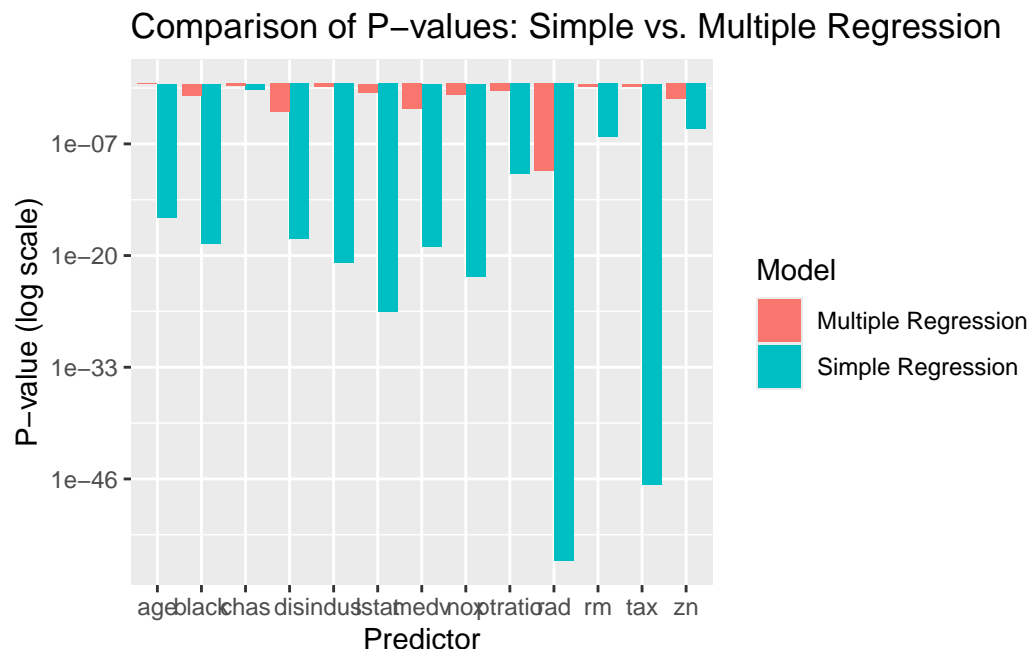

The graph reveals that some predictors retain their significance across both simple and multiple regression models (e.g., rad), while others (e.g., rm, zn, tax) lose significance when other predictors are included. This highlights the importance of using multiple regression to account for overlapping influences among variables, providing a clearer picture of which factors have a direct and unique impact on the crime rate.

2.e.

Polynomial regression model for age

```
model_age <- lm(crim ~ poly(age, 3), data = Boston)
summary(model_age)
```

```
Call:
lm(formula = crim ~ poly(age, 3), data = Boston)

Residuals:
   Min     1Q Median     3Q    Max
-9.762 -2.673 -0.516  0.019 82.842

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)     3.6135     0.3485  10.368  < 2e-16 ***
poly(age, 3)1  68.1820     7.8397   8.697  < 2e-16 ***
poly(age, 3)2  37.4845     7.8397   4.781 2.29e-06 ***
poly(age, 3)3  21.3532     7.8397   2.724  0.00668 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.84 on 502 degrees of freedom
Multiple R-squared:  0.1742,    Adjusted R-squared:  0.1693
F-statistic: 35.31 on 3 and 502 DF,  p-value: < 2.2e-16
```

There is strong evidence of non-linear association between age and crim due to the significance of the quadratic and cubic terms. This suggests that changes in the proportion of older houses have a complex relationship with crime, possibly reflecting different social or economic dynamics across neighborhoods of varying age compositions

Polynomial regression model for tax

```
model_tax <- lm(crim ~ poly(tax, 3), data = Boston)
summary(model_tax)
```

```
Call:
lm(formula = crim ~ poly(tax, 3), data = Boston)

Residuals:
    Min      1Q  Median     3Q     Max
-13.273  -1.389   0.046  0.536  76.950

Coefficients:
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      3.6135     0.3047  11.860  < 2e-16 ***
poly(tax, 3)1 112.6458     6.8537  16.436  < 2e-16 ***
poly(tax, 3)2  32.0873     6.8537   4.682 3.67e-06 ***
poly(tax, 3)3  -7.9968     6.8537  -1.167    0.244
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.854 on 502 degrees of freedom
Multiple R-squared:  0.3689,     Adjusted R-squared:  0.3651
F-statistic:  97.8 on 3 and 502 DF,  p-value: < 2.2e-16
```

There is moderate evidence of non-linear association between tax and crim, primarily driven by the significant quadratic term. This suggests that the relationship between property tax rate and crime is not purely linear—there may be regions where the increase in crime with higher taxes is more or less pronounced. However, since the cubic term is not significant, a quadratic model is likely the best fit for capturing this non-linear trend.

3.a.

Logistic function is given by

$$p(Y = 1|X_1, X_2, X_3) = \frac{e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\beta_3 X_3}}{1 + e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\beta_3 X_3}}$$

Given

$X_1 = 32, X_2 = 3, X_3 = 11$

$\beta_0 = -8, \beta_1 = 0.1, \beta_2 = 1, \beta_3 = -0.04$

Then

$$p(Y = 1|X_1, X_2, X_3) = \frac{e^{-8+0.1*32+1*3+-0.04*11}}{1 + e^{-8+0.1*32+1*3+-0.04*1}} = 0.096$$

The model estimates that the student has approximately a 9.6% chance of receiving an A in the class based on their study hours, GPA, and sleep quality index.

3.b.

Let

$$z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

then

$$p(Y = 1|X_1, X_2, X_3) = \frac{e^z}{1 + e^z}$$

Given

$$p(Y = 1|X_1, X_2, X_3) = 65\% = 0.65$$

$$\frac{e^z}{1 + e^z} = 0.65$$

$$(1 - 0.65)e^z = 0.65$$

$$e^z = \frac{0.65}{0.35} = 1.857$$

$$z = ln(1.857) = 0.619$$

Given

$X_2 = 3, X_3 = 11$

$\beta_0 = -8, \beta_1 = 0.1, \beta_2 = 1, \beta_3 = -0.04$

Then

$$-8 + 0.1(X_1) + 1 * 3 + -0.04 * 11 = 0.619$$

$$X_1 = \frac{0.619 + 8 - 3 + 0.44}{0.1} = 60.59$$

The student would need to study approximately 61 hours to achieve a 65% chance of receiving an A in the class.

3.c.

Let

$$z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

then

$$p(Y = 1 | X_1, X_2, X_3) = \frac{e^z}{1 + e^z}$$

Given

$$p(Y = 1 | X_1, X_2, X_3) = 60\% = 0.6$$

$$\frac{e^z}{1 + e^z} = 0.6$$

$$(1 - 0.6)e^z = 0.6$$

$$e^z = \frac{0.6}{0.4} = 1.5$$

$$z = ln(1.5) = 0.405$$

Given

$X_2 = 3, X_3 = 3$

$\beta_0 = -8, \beta_1 = 0.1, \beta_2 = 1, \beta_3 = -0.04$

Then

$$-8 + 0.1(X_1) + 1 * 3 + -0.04 * 3 = 0.405$$

$$X_1 = \frac{0.405 + 8 - 3 + 0.12}{0.1} = 55.25$$

A student with a GPA of 3.0 and a PSQI score of 3 would need to study approximately 55 hours to have a 60% chance of receiving an A in the class.