

DATA 605: Computational Mathematics Homework 3

Gabriel Campos

Last edited November 21, 2024

Library

```
# Data load
data(cars)
# libraries
library(dplyr)
library(ggplot2)
library(ggthemes)
library(lmtest)
```

Problem 1

Transportation Safety

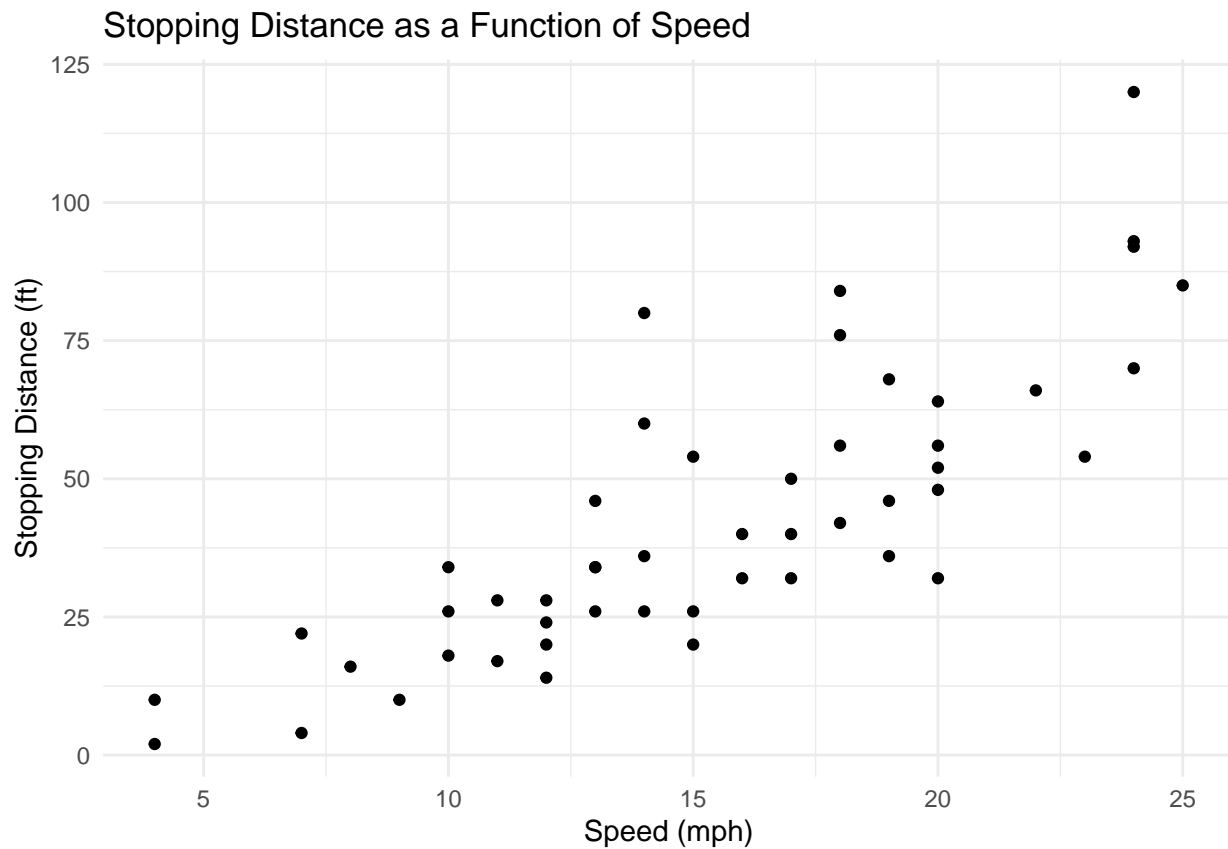
Task

Using the cars dataset in R, perform the following steps:

1 Data Visualization

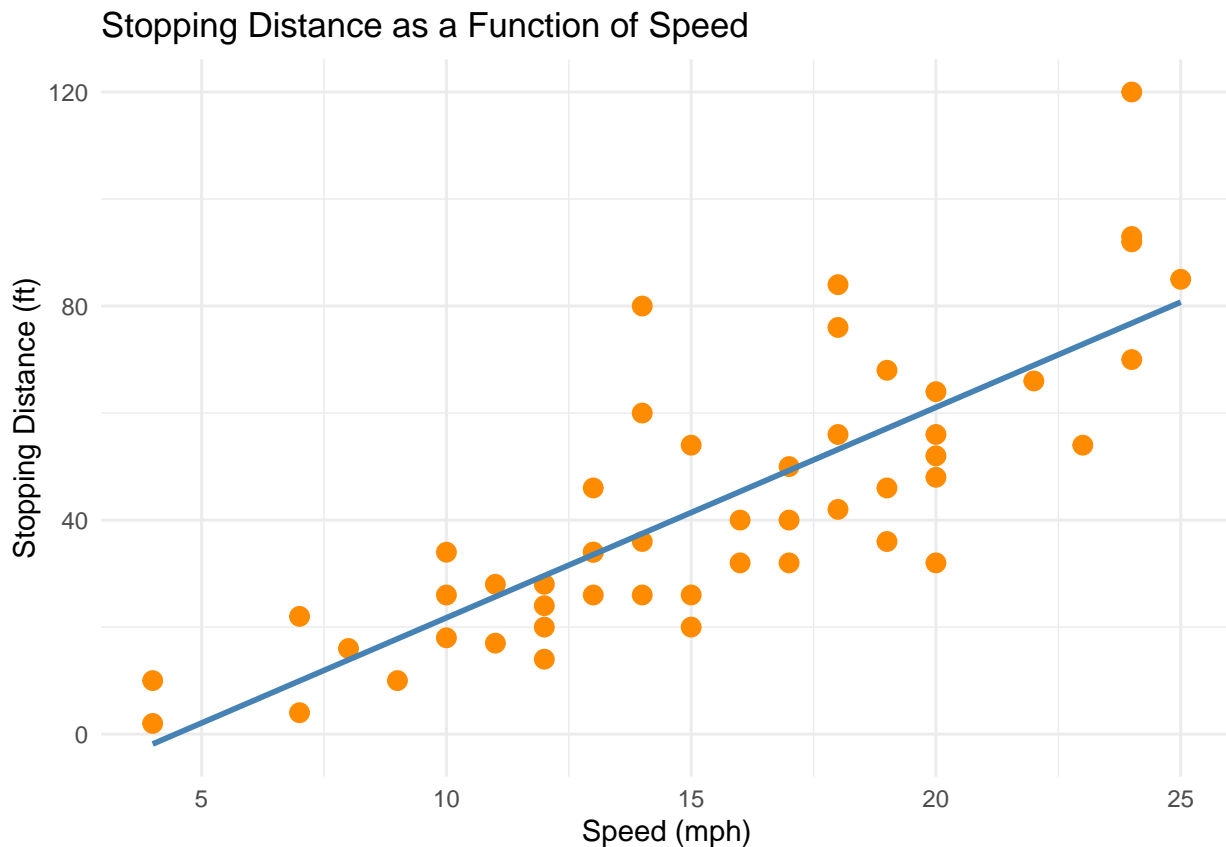
- i Create a scatter plot of stopping distance (dist) as a function of speed (speed).

```
ggplot(cars, aes(x = speed, y = dist)) +
  geom_point() +
  labs(
    title = "Stopping Distance as a Function of Speed",
    x = "Speed (mph)",
    y = "Stopping Distance (ft)"
  ) +
  theme_minimal()
```



ii Add a regression line to the plot to visually assess the relationship.

```
ggplot(cars, aes(x = speed, y = dist)) +
  geom_point(color = "darkorange", size = 3) +
  geom_smooth(method = "lm", color = "steelblue", se = FALSE) +
  labs(
    title = "Stopping Distance as a Function of Speed",
    x = "Speed (mph)",
    y = "Stopping Distance (ft)"
  ) +
  theme_minimal()
```



2 Build a Linear Model

i Construct a simple linear regression model where stopping distance (dist) is the dependent variable and speed (speed) is the independent variable.

```
cars_model <- lm(dist~speed, data = cars)
```

ii Summarize the model to evaluate its coefficients, R-squared value, and p-value.

```
summary(cars_model)
```

```
##
## Call:
## lm(formula = dist ~ speed, data = cars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -29.069  -9.525  -2.272   9.215  43.201
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -17.5791     6.7584  -2.601  0.0123 *
## speed         3.9324     0.4155   9.464 1.49e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.38 on 48 degrees of freedom
```

```
## Multiple R-squared:  0.6511, Adjusted R-squared:  0.6438
## F-statistic: 89.57 on 1 and 48 DF,  p-value: 1.49e-12
```

- **Intercept & Speed** of -17.5791 just indicates the stopping distance at 0 mph. The value is not important only because logically a vehicle not in motion (0mph) does not require stopping distance. The value of 3.9324 indicates for every 1 mph 3.9324 *ft* is needed to stop. Therefore a minimum speed of ≈ 4.48 *mph* is needed before any distance is actually needed to stop.
- **R-Squared** value of 0.6511 indicates $\approx 65\%$ of variance is captured by this relationship, so its a moderate indicator, and there may be several more and potentially stronger factor that can be used to predict the distance needed to stop.
- **P-Value** of $1.49e^{12}$ indicates that the relationship is statistically significant.

3 Model Quality Evaluation

i Calculate and interpret the R-squared value to assess the proportion of variance in stopping distance explained by speed.

- **R-Squared** as indicated above this means $\approx 65\%$ of the variance is captured, making this a moderate factor to stopping distance.
- **R-Squared Adjusted** Frankly, there is only 2 attributes to this table, so adjusting according to the number of predictors of the model makes little difference, therefore the variance of 0.6438 or $\approx 64\%$ is relatively the same to the original, and the remaining factors would include attributes not collected in our dataset.

ii Perform a residual analysis to check the assumptions of the linear regression model, including linearity, homoscedasticity, independence, and normality of residuals.

In order to do this without repeating the steps for Section 4 we leverage the various functions in R through the package **lmtest**

```
cars_resid_values <- resid(cars_model)
cars_fitted_values <- fitted(cars_model)
```

```
summary(cars_resid_values)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -29.069  -9.525  -2.272   0.000   9.215  43.201
```

```
lmtest::bptest(cars_model)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  cars_model
## BP = 3.2149, df = 1, p-value = 0.07297
```

```
cor(cars_fitted_values, cars$dist)
```

```
## [1] 0.8068949
```

```
dwtest(cars_model)
```

```
##
##  Durbin-Watson test
##
## data:  cars_model
## DW = 1.6762, p-value = 0.09522
## alternative hypothesis: true autocorrelation is greater than 0
```

```
ks.test(cars_resid_values, "pnorm", mean=mean(cars_resid_values, sd = sd(cars_resid_values)))
```

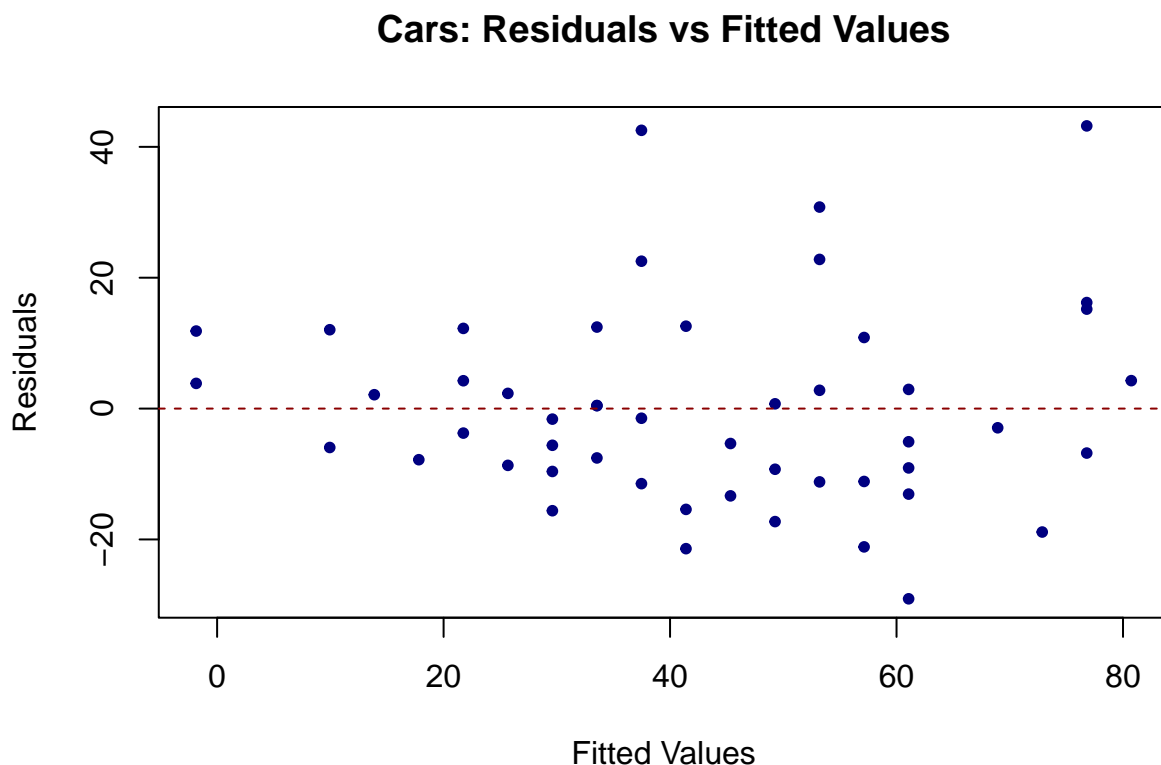
```
##
## Asymptotic one-sample Kolmogorov-Smirnov test
##
## data: cars_resid_values
## D = 0.49833, p-value = 3.283e-11
## alternative hypothesis: two-sided
```

- We have an mean of 0 which is ideal
- A correlation of 0.8068949 suggests strong linearity since it is close to 1.
- the p – value > 0.05 through the *Breusch-Pagan test* (`bptest()`) which it is at 0.7297 indicates homoscedasticity.
- the p – value through the *Durbin-Watson test* (`dwtest()`) of 0.09522 indicated residuals are independent or has no significant autocorrelation in residuals.
- the p – value through the *Kolmogorav-Smirnov test* (`ks.test()`) of 3.283×10^{-11} / indicates its not normally distributed.

4 Residual Analysis

i Plot the residuals versus fitted values to check for any patterns.

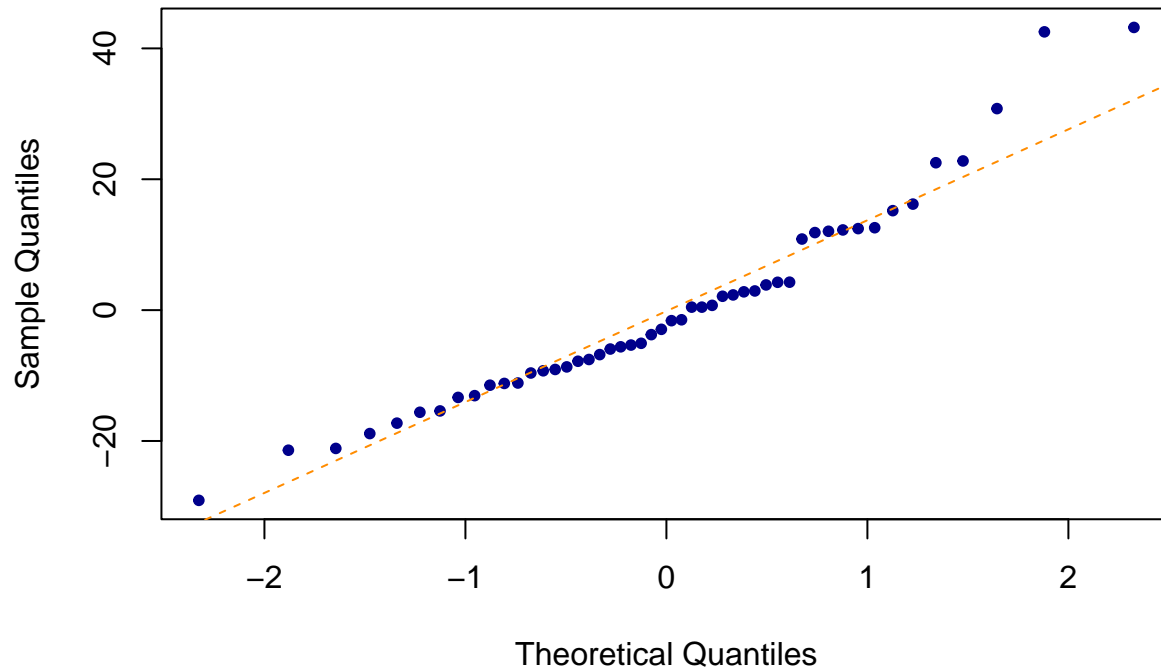
```
plot(cars_fitted_values, cars_resid_values,
     xlab = "Fitted Values",
     ylab = "Residuals",
     main = "Cars: Residuals vs Fitted Values",
     pch = 20,
     col = "navyblue")
abline(h = 0, col = "darkred", lty = 2)
```



ii Create a Q-Q plot of the residuals to assess normality.

```
qqnorm(cars_resid_values,  
       main = "Q-Q Plot of Residuals",  
       pch = 20,  
       col = "darkblue")  
qqline(cars_resid_values, col = "darkorange", lty = 2)
```

Q-Q Plot of Residuals



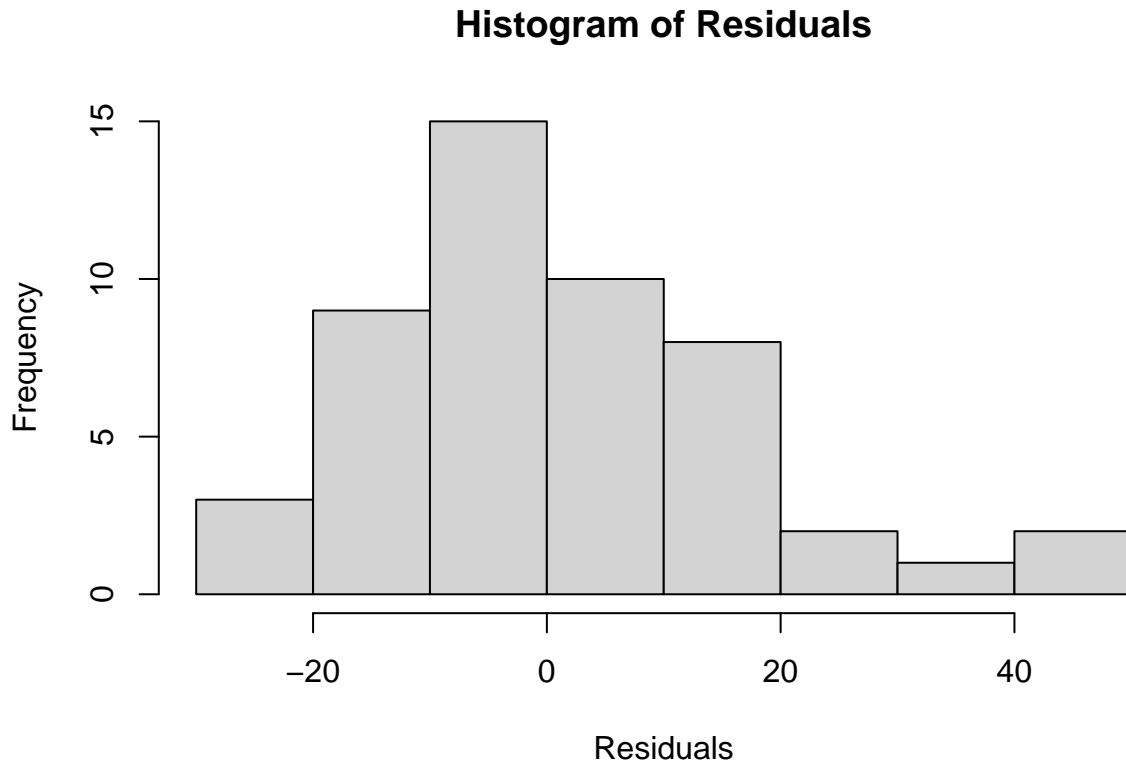
iii Perform a Shapiro-Wilk test for normality of residuals.

```
cars_shapiro_test <- shapiro.test(cars_resid_values)  
cars_shapiro_test
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: cars_resid_values  
## W = 0.94509, p-value = 0.02152
```

iv Plot a histogram of residuals to further check for normality.

```
hist(cars_resid_values,  
     breaks = 10, # Adjust the number of bins as needed  
     col = "lightgrey",  
     main = "Histogram of Residuals",  
     xlab = "Residuals",  
     ylab = "Frequency",  
     border = "black")
```



Conclusion Based on the model summary and residual analysis, determine whether the linear model is appropriate for this data. Discuss any potential violations of model assumptions and suggest improvements if necessary.

The model for this data set is reasonably appropriate, as it demonstrates moderate explanatory power ($R^2 \approx 0.65$). While the Q-Q plot and histogram suggest residuals indicate near-normal residuals, the Shapiro-Wilk test ($p = 0.02152$) suggests mild non-normality. Additionally, the residuals vs fitted plot reveals 3 outliers that may impact the models validity. To improve the model, we can consider removing or adjusting for these outliers or testing a non-linear regression model to capture potential curvature in the data.

Problem 2

Health Policy Analyst

As a health policy analyst for an international organization, you are tasked with analyzing data from the World Health Organization (WHO) to inform global health policies. The dataset provided (who.csv) contains crucial health indicators for various countries from the year 2008. The variables include:

- Country: Name of the country
- LifeExp: Average life expectancy for the country in years
- InfantSurvival: Proportion of those surviving to one year or more
- Under5Survival: Proportion of those surviving to five years or more
- TBFree: Proportion of the population without TB
- PropMD: Proportion of the population who are MDs
- PropRN: Proportion of the population who are RNs
- PersExp: Mean personal expenditures on healthcare in US dollars at average exchange rate
- GovtExp: Mean government expenditures per capita on healthcare, US dollars at average exchange rate
- TotExp: Sum of personal and government expenditures

Your analysis will directly influence recommendations for improving global life expectancy and the allocation of healthcare resources.

```
who_df <- read.csv("who.csv", fileEncoding = "UTF-8")%>%
  select(-LifeExp.1,-X)
```

Question 1

Initial Assessment of Healthcare Expenditures and Life Expectancy

Task Create a scatterplot of LifeExp vs. TotExp to visualize the relationship between healthcare expenditures and life expectancy across countries. Then, run a simple linear regression with LifeExp as the dependent variable and TotExp as the independent variable (without transforming the variables).

- Provide and interpret the F-statistic, R-squared value, standard error, and p-values.
- Discuss whether the assumptions of simple linear regression (linearity, independence, homoscedasticity, and normality of residuals) are met in this analysis.

```
str(who_df)
```

```
## 'data.frame': 42 obs. of 10 variables:
## $ Country : chr "Afghanistan" "Albania" "Algeria" "Andorra" ...
## $ LifeExp : int 42 71 71 82 41 73 75 69 82 80 ...
## $ InfantSurvival: num 0.835 0.985 0.967 0.997 0.846 0.99 0.986 0.979 0.995 0.996 ...
## $ Under5Survival: num 0.743 0.983 0.962 0.996 0.74 0.989 0.983 0.976 0.994 0.996 ...
## $ TBFree : num 0.998 1 0.999 1 0.997 ...
## $ PropMD : num 2.29e-04 1.14e-03 1.06e-03 3.30e-03 7.04e-05 ...
## $ PropRN : num 0.000572 0.004614 0.002091 0.0035 0.001146 ...
## $ PersExp : int 20 169 108 2589 36 503 484 88 3181 3788 ...
## $ GovtExp : int 92 3128 5184 169725 1620 12543 19170 1856 187616 189354 ...
## $ TotExp : int 112 3297 5292 172314 1656 13046 19654 1944 190797 193142 ...
```

```
summary(who_df)
```

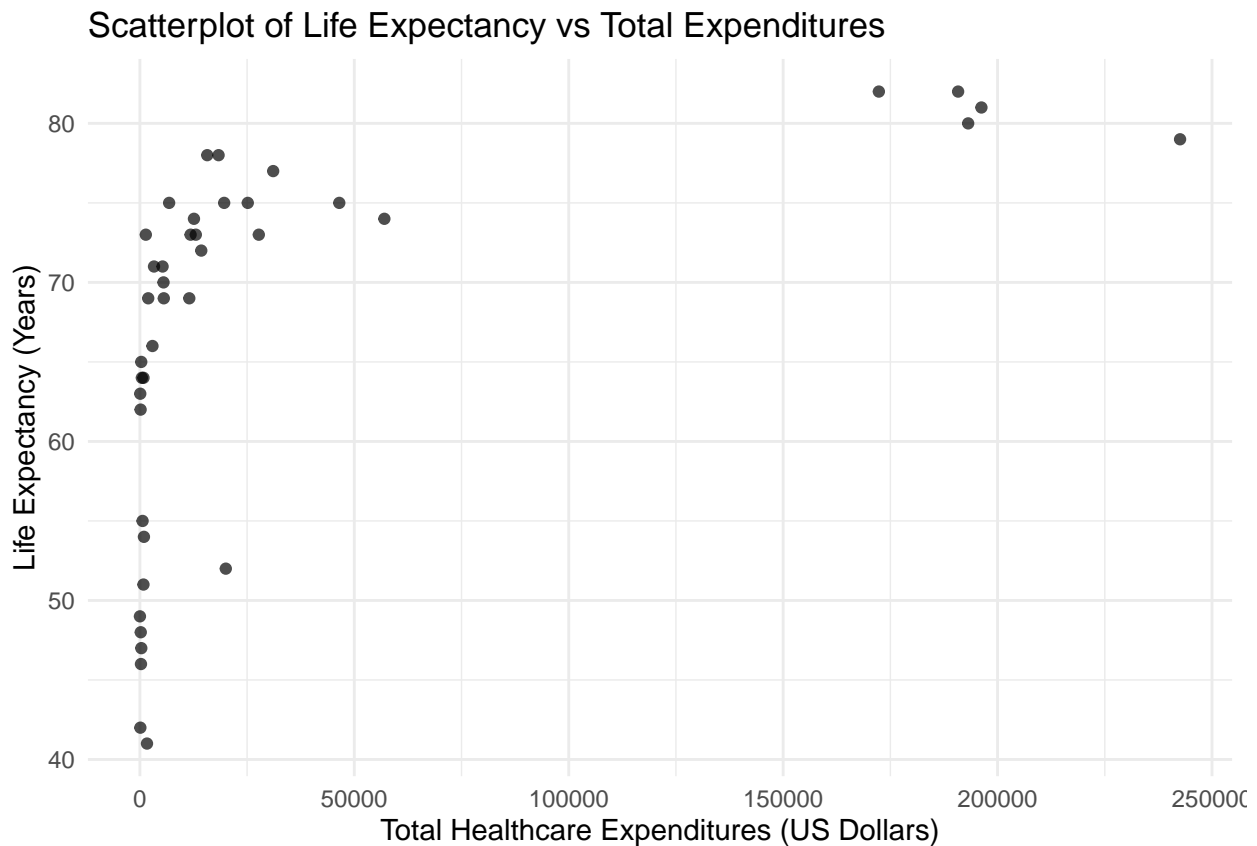
```
## Country LifeExp InfantSurvival Under5Survival
## Length:42 Min. :41.00 Min. :0.8350 Min. :0.7400
## Class :character 1st Qu.:62.00 1st Qu.:0.9270 1st Qu.:0.9110
## Mode :character Median :71.00 Median :0.9810 Median :0.9790
## Mean :66.76 Mean :0.9561 Mean :0.9357
## 3rd Qu.:75.00 3rd Qu.:0.9900 3rd Qu.:0.9880
## Max. :82.00 Max. :0.9970 Max. :0.9960
## NA's :1 NA's :1 NA's :1
## TBFree PropMD PropRN PersExp
## Min. :0.9929 Min. :0.0000245 Min. :0.0001649 Min. : 3.0
## 1st Qu.:0.9973 1st Qu.:0.0001719 1st Qu.:0.0007410 1st Qu.: 36.0
## Median :0.9993 Median :0.0010471 Median :0.0019340 Median : 198.0
## Mean :0.9982 Mean :0.0012641 Mean :0.0032899 Mean : 603.5
## 3rd Qu.:0.9997 3rd Qu.:0.0014286 3rd Qu.:0.0046694 3rd Qu.: 484.0
## Max. :1.0000 Max. :0.0047587 Max. :0.0140792 Max. :3788.0
## NA's :1 NA's :1 NA's :1 NA's :1
## GovtExp TotExp
## Min. : 10 Min. : 13
## 1st Qu.: 780 1st Qu.: 833
## Median : 5394 Median : 5574
## Mean : 32547 Mean : 33150
## 3rd Qu.: 19604 3rd Qu.: 20035
```



```
## Max.      :239105    Max.      :242556
## NA's      :1         NA's      :1
```

Create a scatterplot of LifeExp vs. TotExp

```
ggplot(who_df, aes(x = TotExp, y = LifeExp)) +
  geom_point(alpha = 0.7) +
  labs(
    title = "Scatterplot of Life Expectancy vs Total Expenditures",
    x = "Total Healthcare Expenditures (US Dollars)",
    y = "Life Expectancy (Years)"
  ) +
  theme_minimal()
```

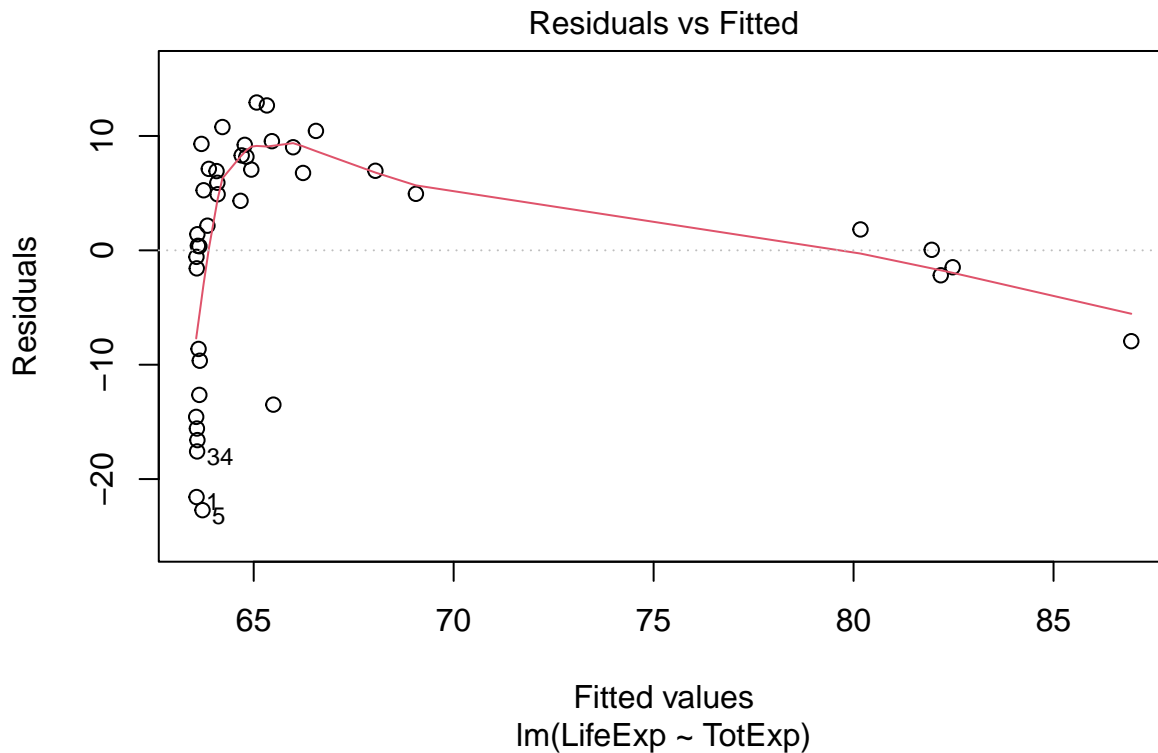


```
who_model <- lm(LifeExp ~ TotExp, data = who_df)
```

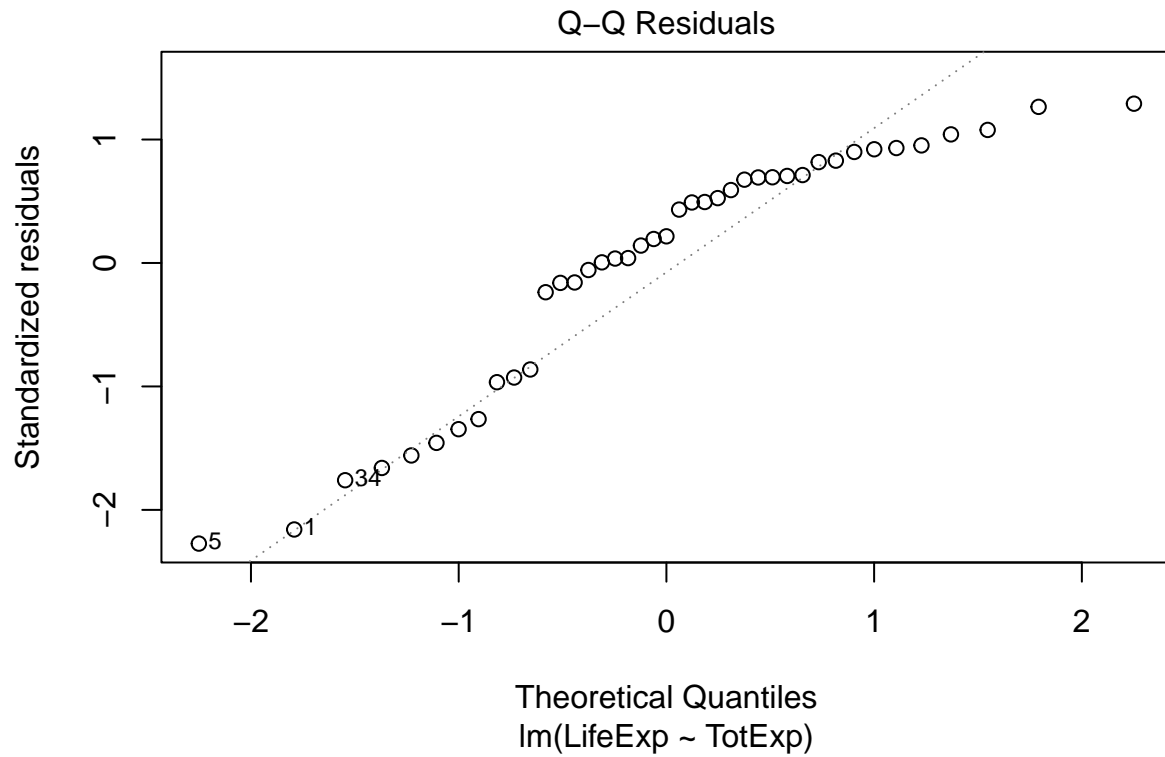
```
summary(who_model)
```

```
##
## Call:
## lm(formula = LifeExp ~ TotExp, data = who_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.720  -7.944   2.157   7.122  12.926
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.356e+01  1.788e+00  35.553  < 2e-16 ***
```

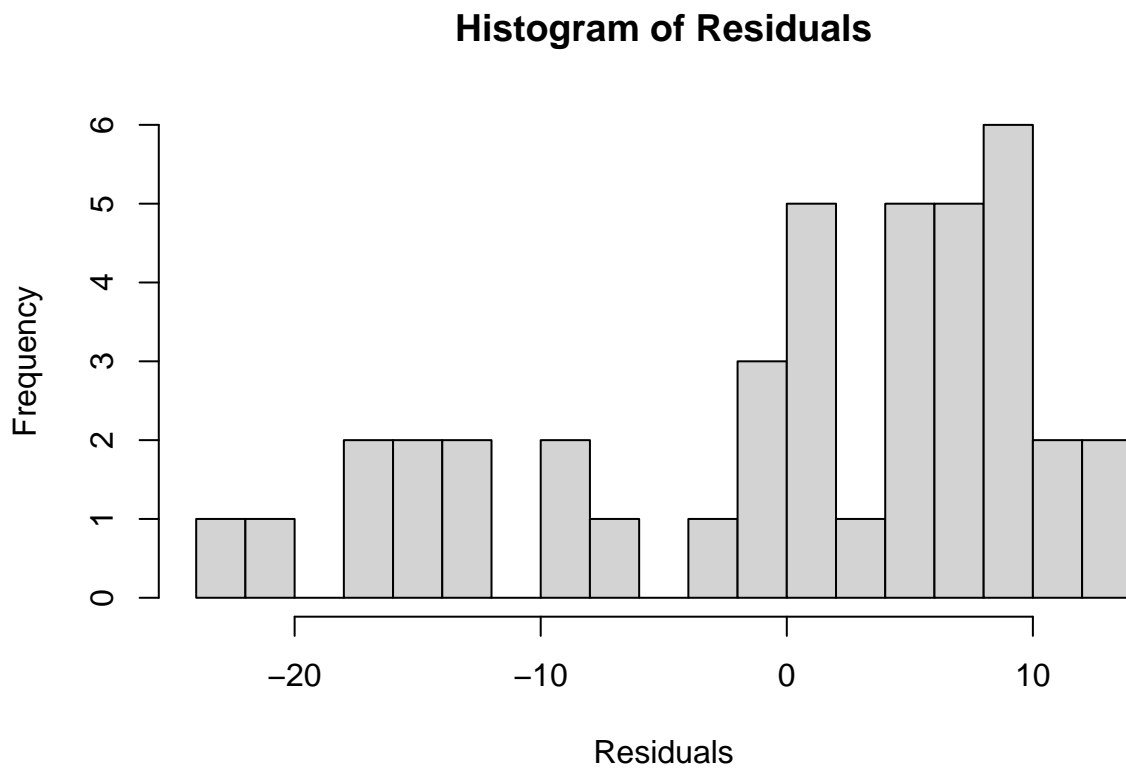
```
## TotExp      9.641e-05  2.493e-05   3.867 0.000406 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.15 on 39 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.2772, Adjusted R-squared:  0.2587
## F-statistic: 14.96 on 1 and 39 DF,  p-value: 0.0004064
#Residuals vs Fitted
plot(who_model, which = 1)
```



```
# Q-Q plot
plot(who_model, which = 2)
```



```
# Histogram of residuals
hist(residuals(who_model), breaks = 20, main = "Histogram of Residuals", xlab = "Residuals")
```



Discussion

i. Consider the implications of your findings for health policy. Are higher healthcare expenditures generally associated with longer life expectancy?

The data shows a moderate positive relationship between healthcare expenditures and life expectancy.

- F -statistic of 14.96 and p -value ≈ 0.0004 indicates statistical significance but explains less variability.
- Standard Error (SE) of 10.15 a relatively good fit with the precision of the estimated regression coefficient (like the slope or intercept).
- Positive coefficient 9.641×10^{-5} indicates as healthcare expenditure increases so does life expectancy.
- The plots show no normality with the residuals.
- There are no signs of homoscedasticity as the points do not show randomness, a funnel or fan-like shape.
- The pattern with these two variables indicates there is no independence.
- The Q-Q plot points deviate significantly indicating Non-Normality, which is supported by the Histogram of the Residuals.
- The relationship between the variables does not seem linear.

ii. What do the assumptions of the regression model suggest about the reliability of this relationship? $R^2 = 0.2772$ suggests the relationship is not particularly strong, implying that while healthcare expenditures explain some variation in life expectancy, the model does not capture all the influencing factors.

Question 2

Transforming Variables for a Better Fit

Task Recognizing potential non-linear relationships, transform the variables as follows

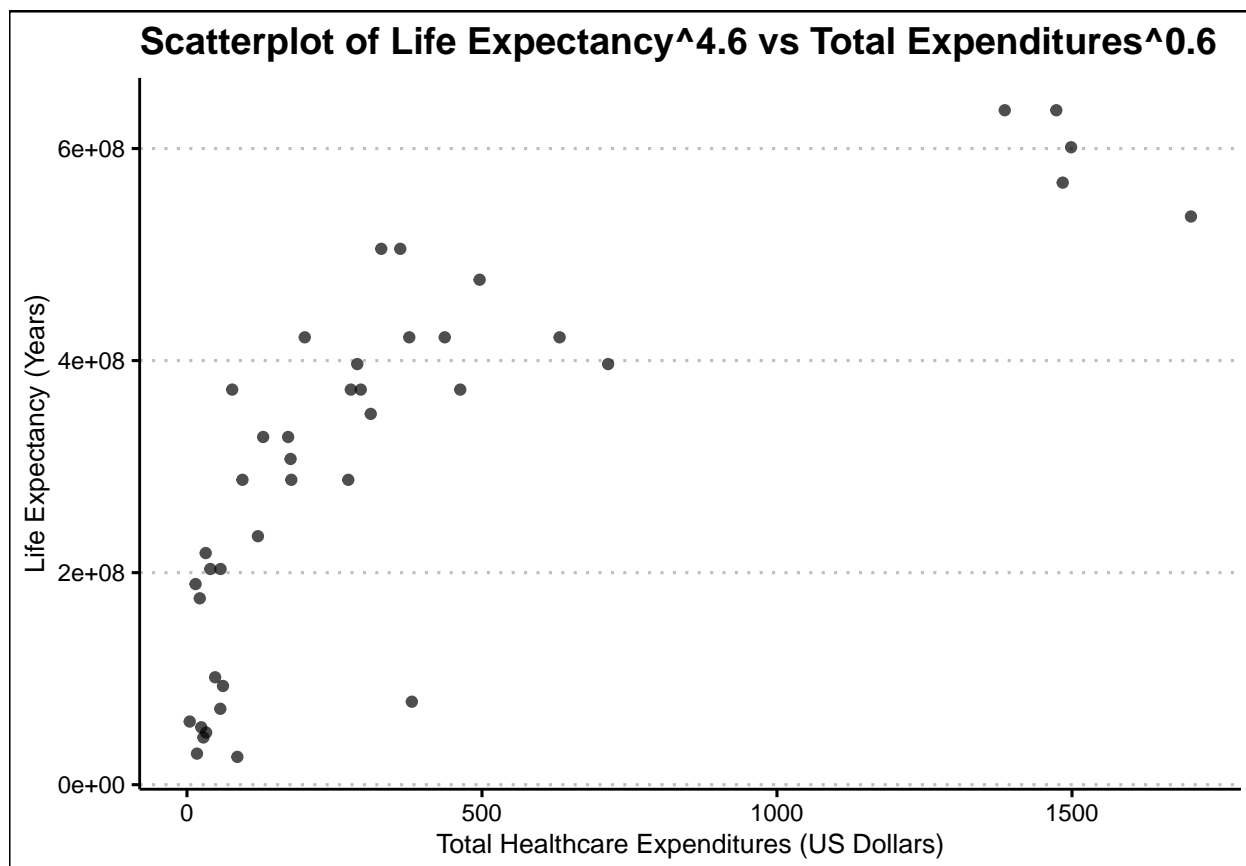
- Raise life expectancy to the 4.6 power ($\text{LifeExp}^{4.6}$).
- Raise total expenditures to the 0.06 power ($\text{TotExp}^{0.06}$), which is nearly a logarithmic transformation.

Create a new scatterplot with the transformed variables and re-run the simple linear regression model.

- Provide and interpret the F-statistic, R-squared value, standard error, and p-values for the transformed model.
- Compare this model to the original model (from Question 1). Which model provides a better fit, and why?

```
who_df$LifeExp_t <- who_df$LifeExp^4.6
who_df$TotExp_t <- who_df$TotExp^0.06

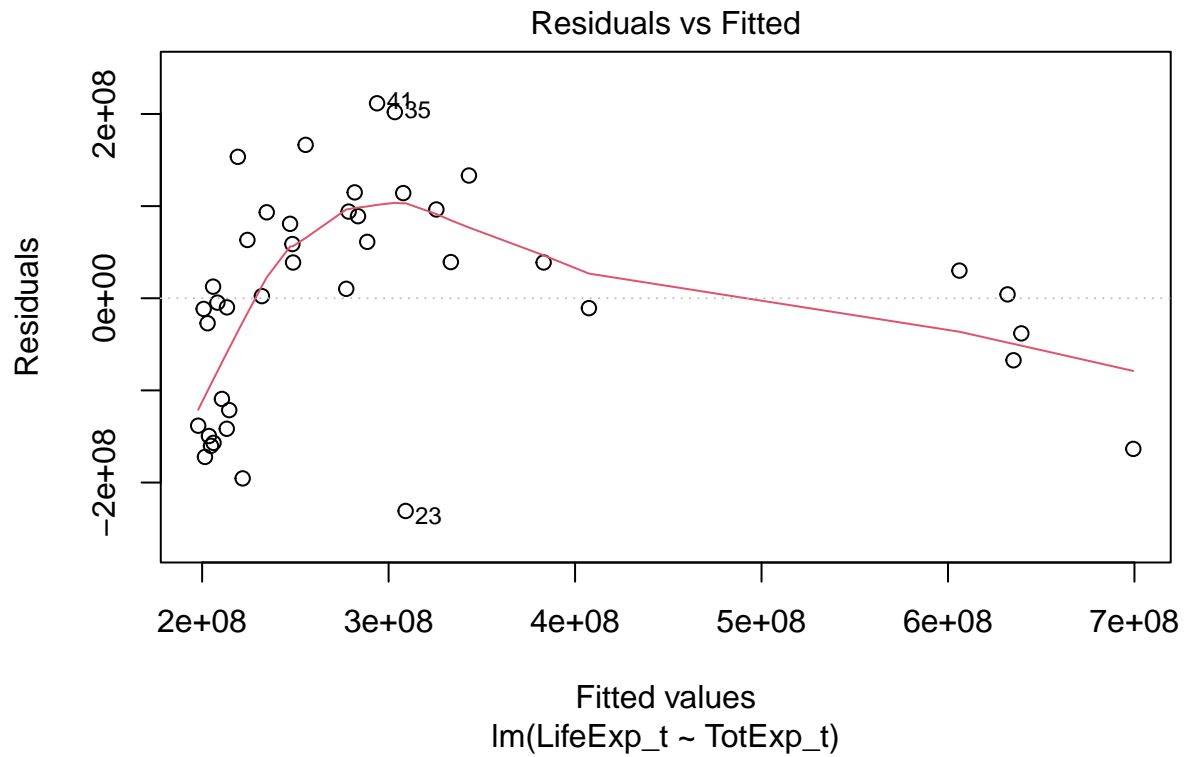
ggplot(who_df, aes(x = TotExp_t, y = LifeExp_t)) +
  geom_point(alpha = 0.7) +
  labs(
    title = "Scatterplot of Life Expectancy^4.6 vs Total Expenditures^0.06",
    x = "Total Healthcare Expenditures (US Dollars)",
    y = "Life Expectancy (Years)"
  ) +
  theme_clean()
```



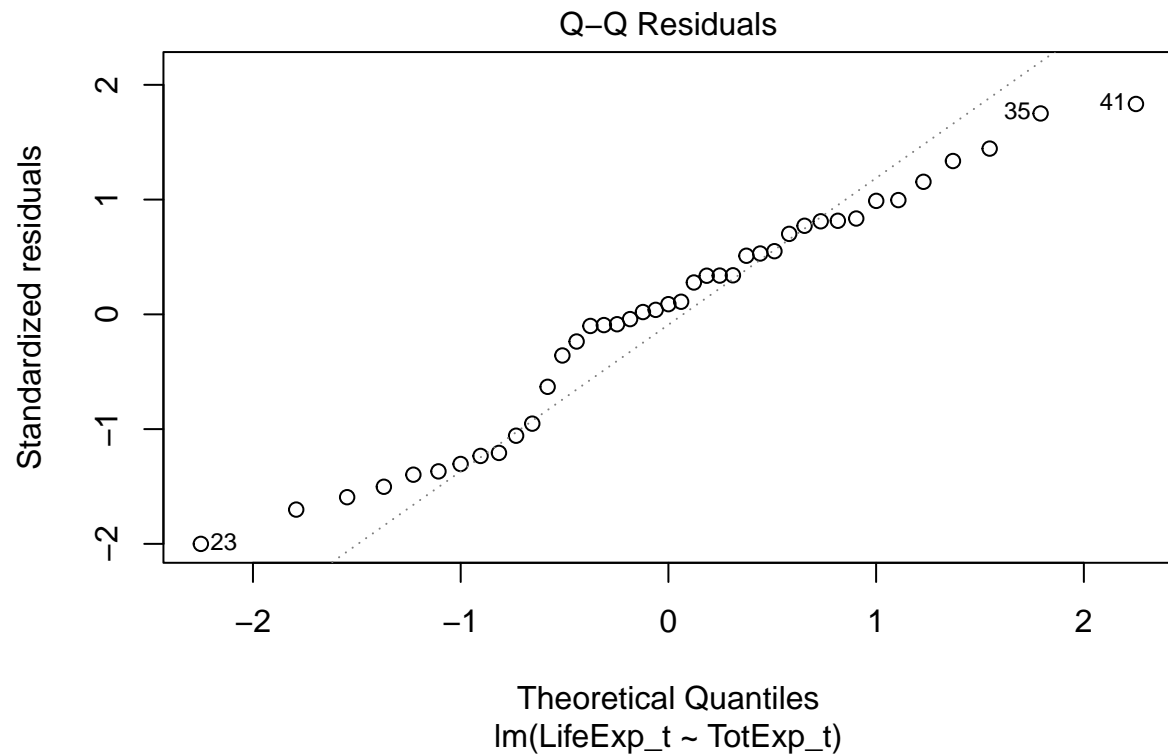
```
who_model_t <- lm(LifeExp_t ~ TotExp_t, data = who_df)
summary(who_model_t)
```

```
##
## Call:
## lm(formula = LifeExp_t ~ TotExp_t, data = who_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -230855907 -109289555  10242350   89067282  211587690
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 196498750   23177824   8.478 2.21e-10 ***
## TotExp_t      295518      39460    7.489 4.61e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 116900000 on 39 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.5898, Adjusted R-squared:  0.5793
## F-statistic: 56.09 on 1 and 39 DF,  p-value: 4.61e-09

#Residuals vs Fitted
plot(who_model_t, which = 1)
```

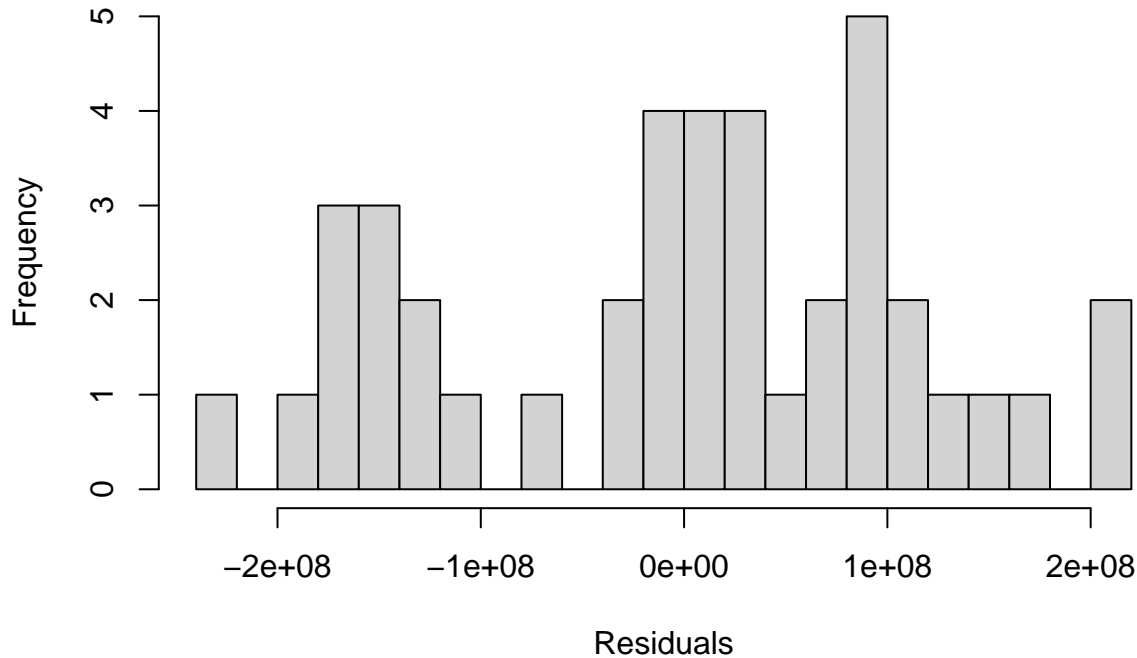


```
# Q-Q plot
plot(who_model_t, which = 2)
```



```
# Histogram of residuals
hist(residuals(who_model_t), breaks = 20, main = "Histogram of Residuals", xlab = "Residuals")
```

Histogram of Residuals



- F – statistic of 56.09 and p – value = 4.61×10^9 indicates greater statistical significance and captures greater proportion of variability.
- Standard Error (SE) of 116,900,000 can be viewed as a worse fit with the precision of the estimated regression coefficient but is more likely a result of the greater magnitude.
- Positive coefficient 295518 indicates as healthcare expenditure increases so does life expectancy still, but to a greater magnitude.
- The plots show left skewed results with a more normalized residual.
- There are no signs homoscedasticity as the points do not show randomness, a funnel or fan like shape.
- The pattern with these two variable indicates there is no independence.
- The Q-Q plot points still deviates albeit no as significantly, indicating Normality, which is supported by the Histogram of the Residuals.
- The relationship between the variables does seem more linear when compared to its predecessor, but still non-linear overall.

Discussion

i How do the transformations impact the interpretation of the relationship between healthcare spending and life expectancy?

$R^2 = 0.5898$ suggests the relationship is strong, particularly when compare to the predecessor.

ii Why might the transformed model be more appropriate for policy recommendations?

The stronger relationship is highlighted after the transformation, which better accounts for the complex, non-linear relationship between healthcare expenditures and life expectancy. This approach reduces risk in model specification and captures the diminishing return on healthcare spending, that has less of an impact after a certain threshold.

Question 3

Forecasting Life Expectancy Based on Transformed Expenditures

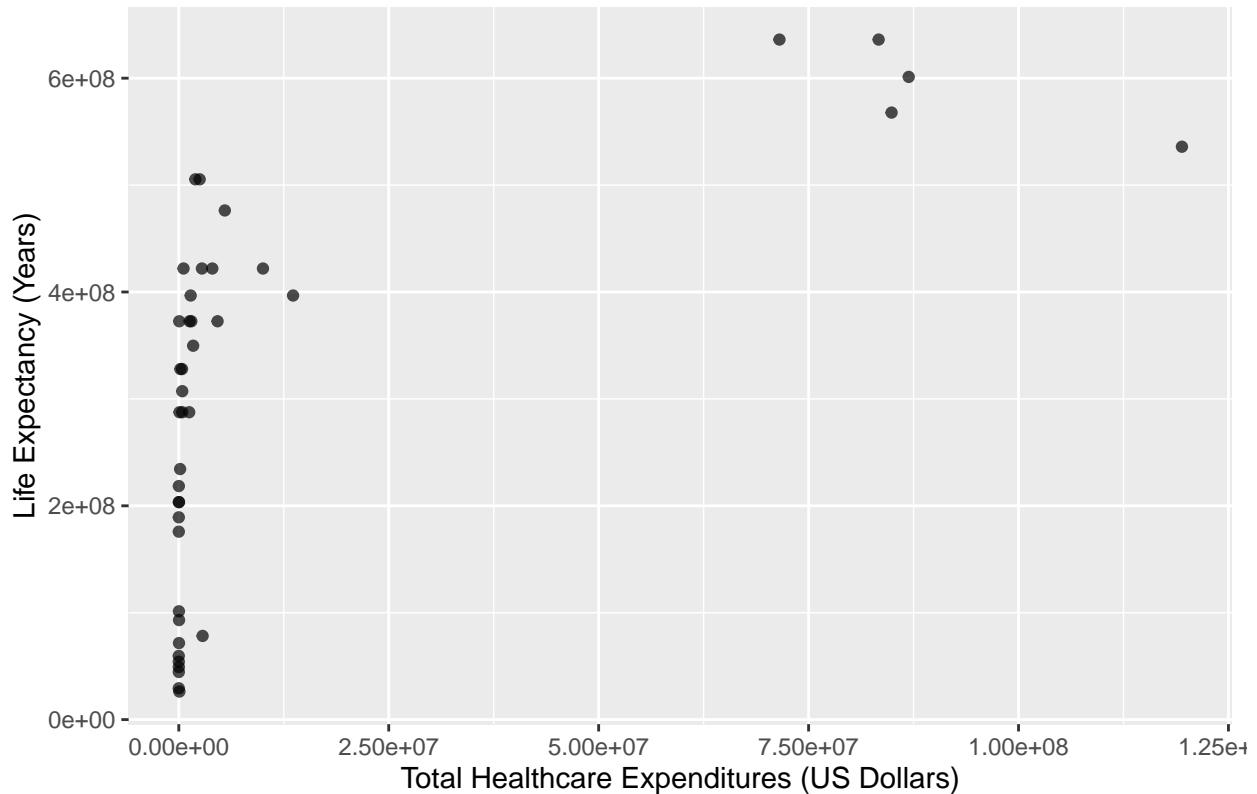
Task Using the results from the transformed model in Question 2, forecast the life expectancy for countries with the following transformed total expenditures ($TotalExp^{0.06}$):

- When $TotalExp^{0.06} = 1.5$
- When $TotalExp^{0.06} = 2.5$

```
who_df$TotExp_t15 <- who_df$TotExp^1.5
who_df$TotExp_t25 <- who_df$TotExp^2.5
```

```
ggplot(who_df, aes(x = TotExp_t15, y = LifeExp_t)) +
  geom_point(alpha = 0.7) +
  labs(
    title = "Scatterplot of Life Expectancy^1.5 vs Total Expenditures^0.6",
    x = "Total Healthcare Expenditures (US Dollars)",
    y = "Life Expectancy (Years)"
  )
```

Scatterplot of Life Expectancy^{1.5} vs Total Expenditures^{0.6}



```
who_model_t15 <- lm(LifeExp_t ~ TotExp_t15, data = who_df)
summary(who_model_t15)
```

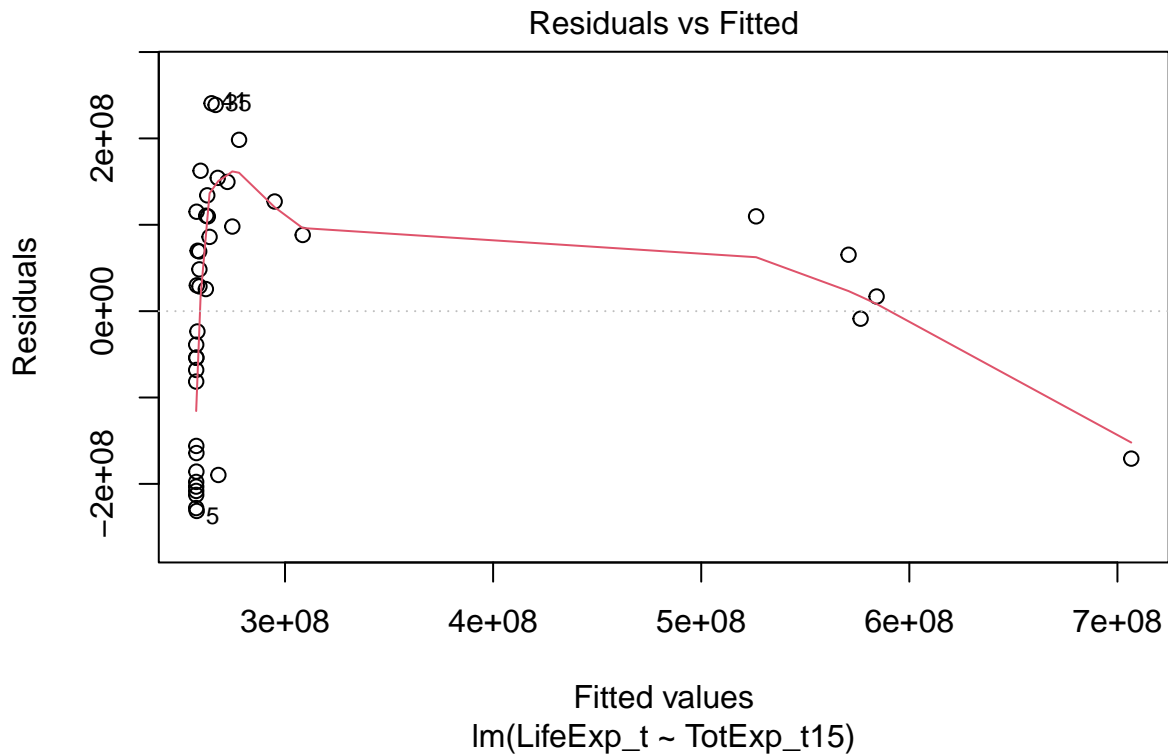
```
##
## Call:
## lm(formula = LifeExp_t ~ TotExp_t15, data = who_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -231330911 -156053328  28671772 109763522 240690290
##
## Coefficients:
```



```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.573e+08  2.423e+07  10.618 4.55e-13 ***
## TotExp_t15  3.761e+00  7.618e-01   4.938 1.53e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 143100000 on 39 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.3847, Adjusted R-squared:  0.3689
## F-statistic: 24.38 on 1 and 39 DF,  p-value: 1.525e-05
```

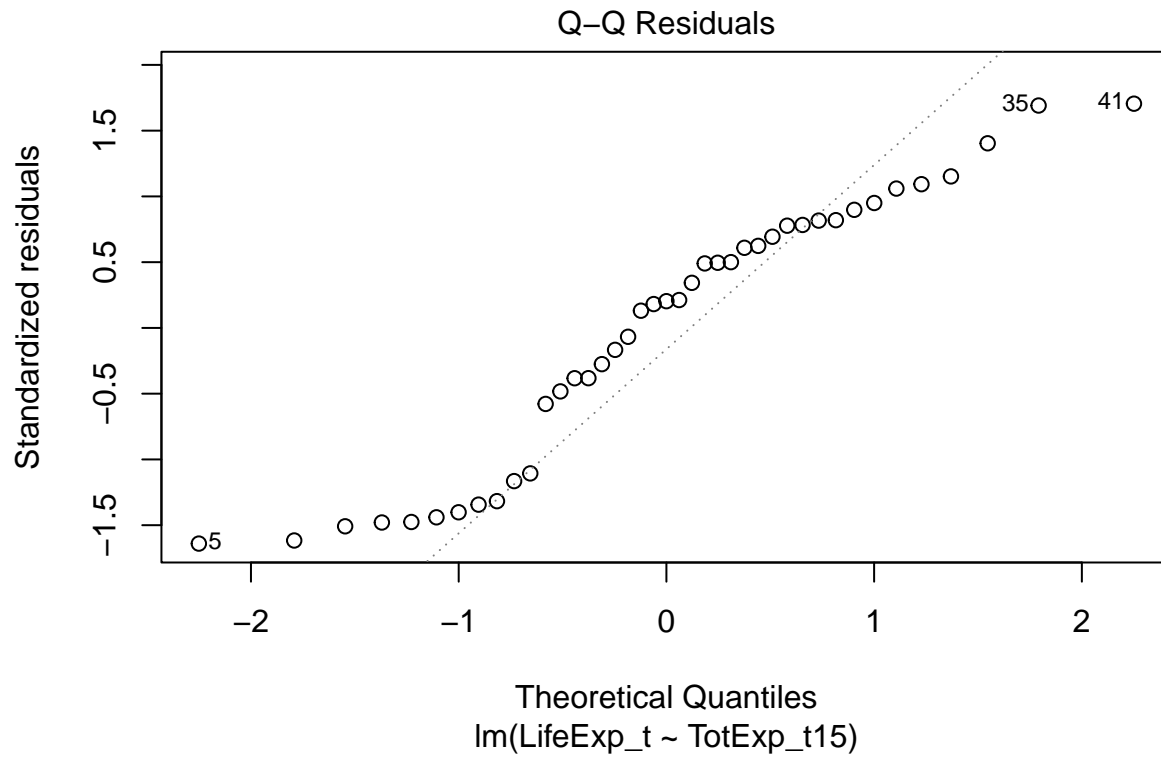
```
#Residuals vs Fitted
```

```
plot(who_model_t15, which = 1)
```

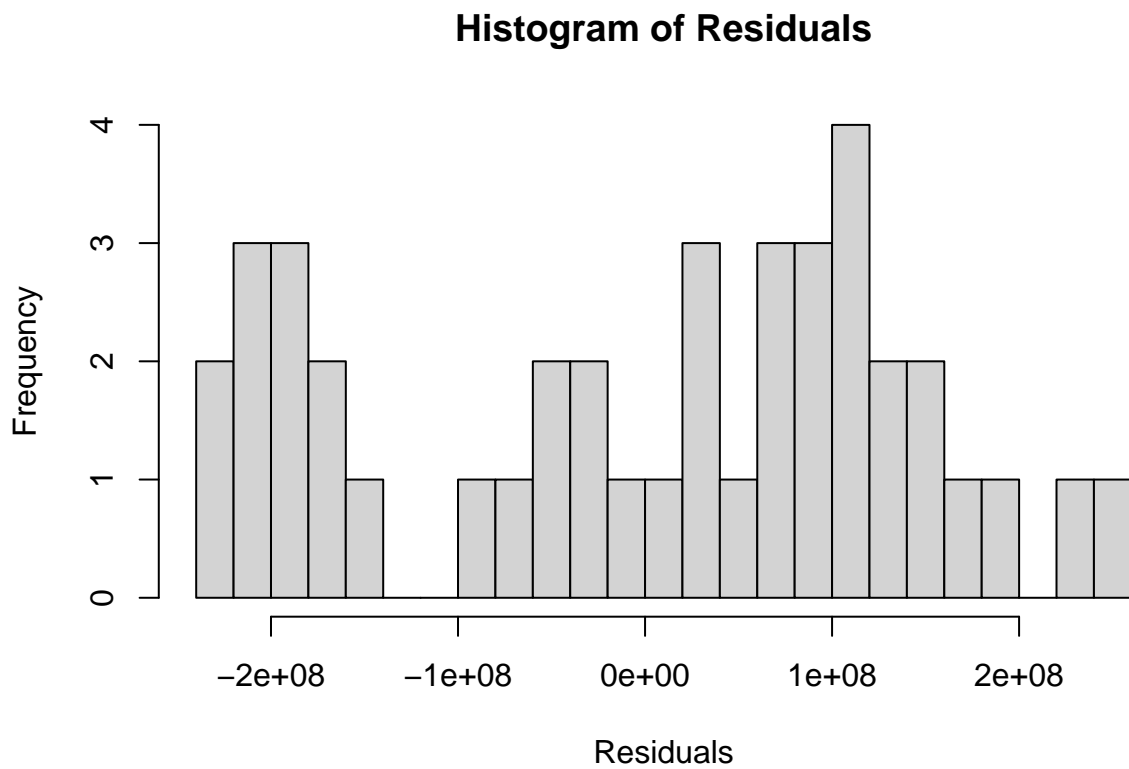


```
# Q-Q plot
```

```
plot(who_model_t15, which = 2)
```



```
# Histogram of residuals
hist(residuals(who_model_t15), breaks = 20, main = "Histogram of Residuals", xlab = "Residuals")
```

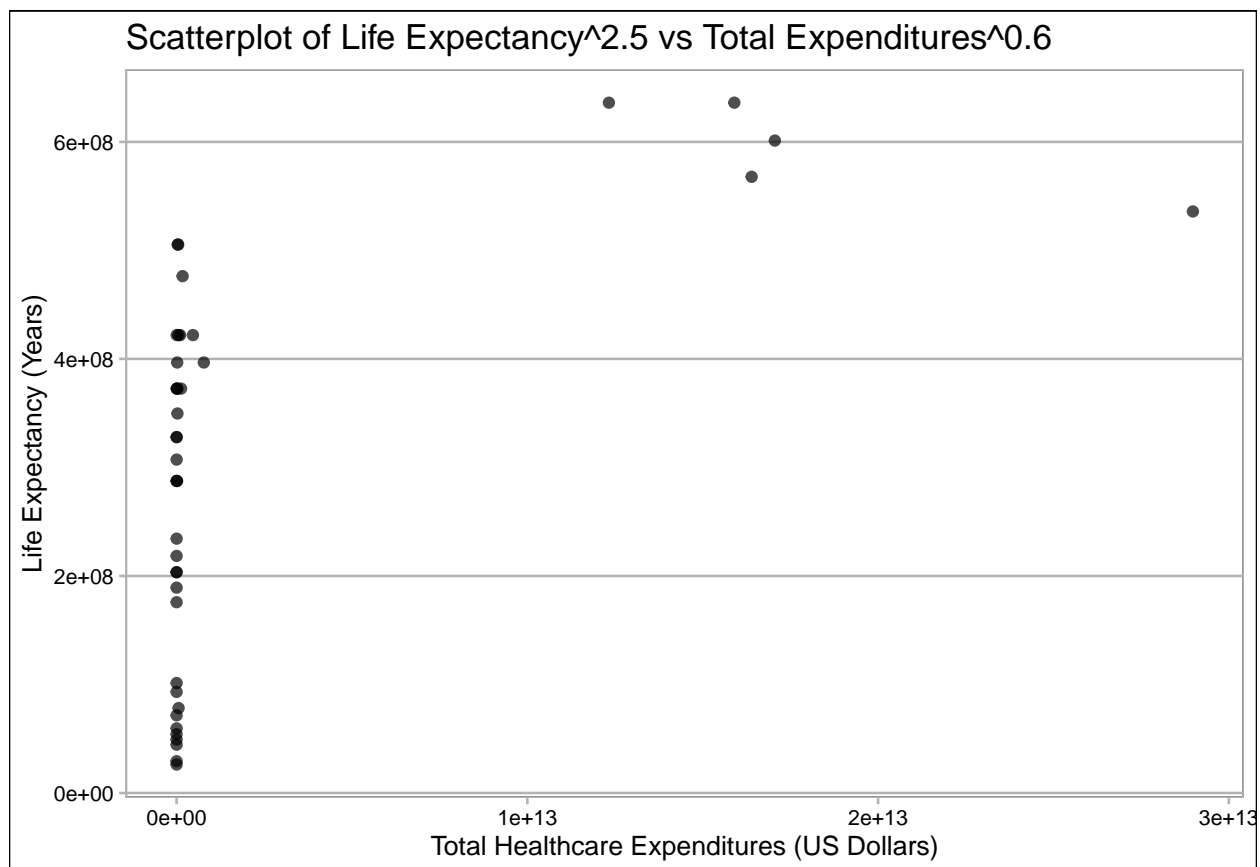


```
ggplot(who_df, aes(x = TotExp_t25, y = LifeExp_t)) +
  geom_point(alpha = 0.7) +
  labs(
```

```

title = "Scatterplot of Life Expectancy^2.5 vs Total Expenditures^0.6",
x = "Total Healthcare Expenditures (US Dollars)",
y = "Life Expectancy (Years)"
) +
theme_calc()

```



```

who_model_t25 <- lm(LifeExp_t ~ TotExp_t25, data = who_df)
summary(who_model_t25)

```

```

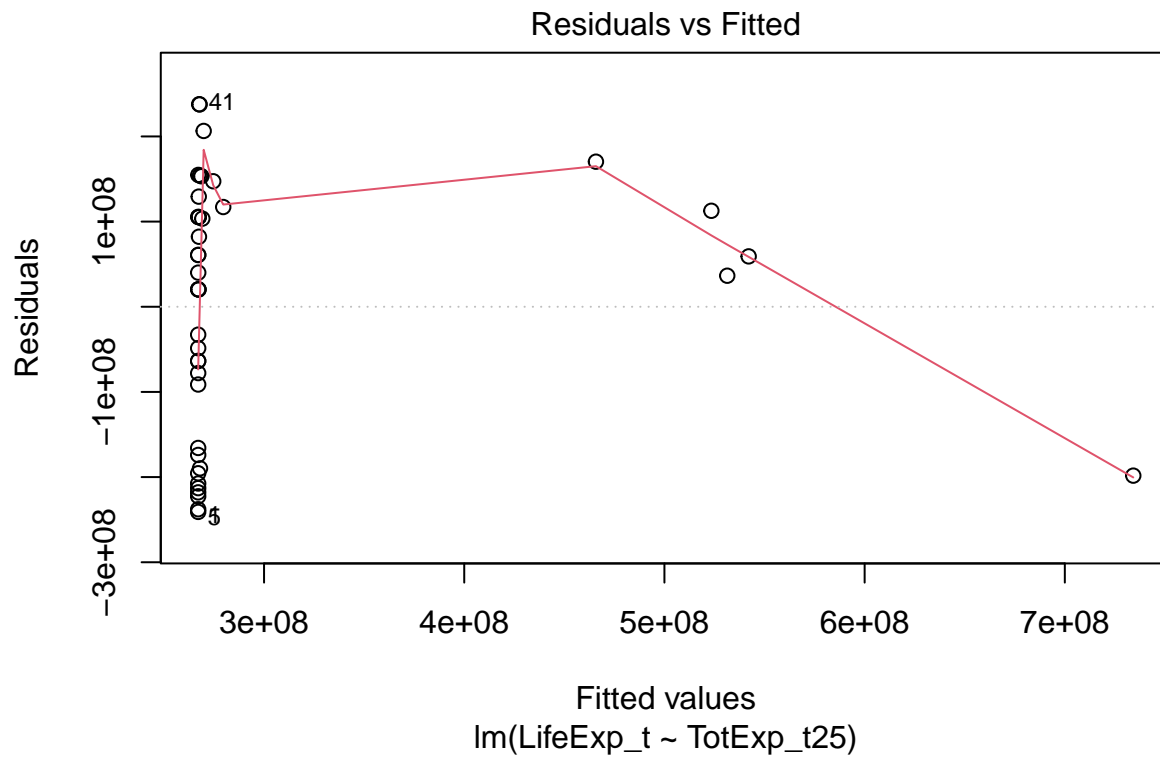
##
## Call:
## lm(formula = LifeExp_t ~ TotExp_t25, data = who_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -240868624 -165783689   36440609  112692262  237804522
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.671e+08  2.500e+07  10.683 3.81e-13 ***
## TotExp_t25   1.612e-05  3.768e-06   4.278 0.000118 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 150500000 on 39 degrees of freedom
## (1 observation deleted due to missingness)

```

```
## Multiple R-squared:  0.3194, Adjusted R-squared:  0.302
## F-statistic:  18.3 on 1 and 39 DF,  p-value: 0.0001182
```

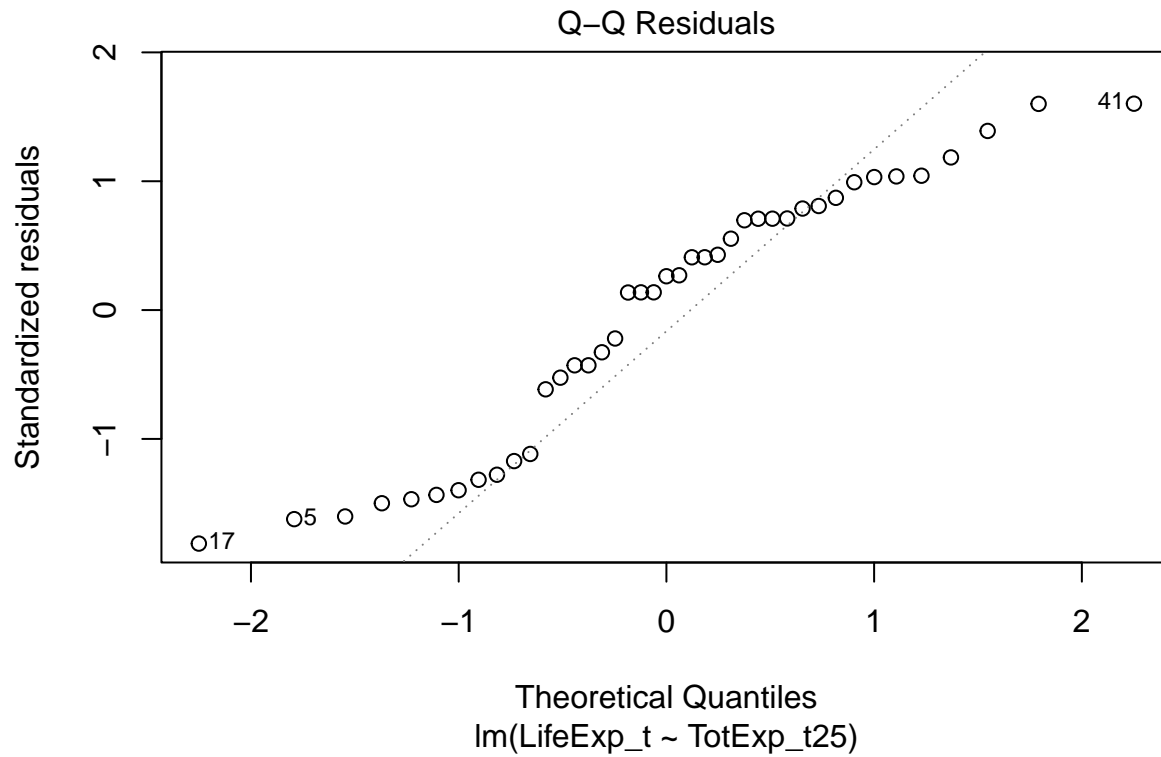
```
#Residuals vs Fitted
```

```
plot(who_model_t25, which = 1)
```

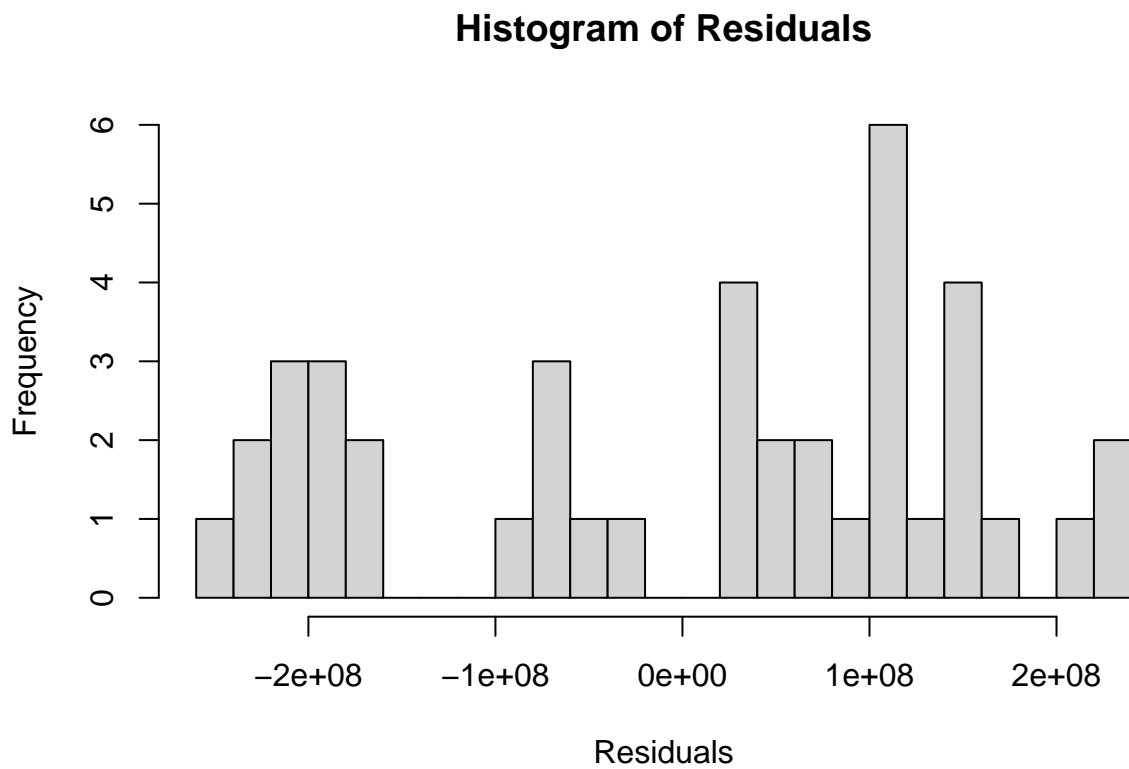


```
# Q-Q plot
```

```
plot(who_model_t25, which = 2)
```



```
# Histogram of residuals
hist(residuals(who_model_t25), breaks = 20, main = "Histogram of Residuals", xlab = "Residuals")
```



Discussion: Discuss the implications of these forecasts for countries with different levels of healthcare spending. What do these predictions suggest about the potential impact of increasing healthcare expenditures

on life expectancy?

Question 4

Interaction Effects in Multiple Regression

Task Build a multiple regression model to investigate the combined effect of the proportion of MDs and total healthcare expenditures on life expectancy. Specifically, use the model:

$$LifeExp = b_0 + b_1 \times PropMD + b_2 \times TotExp + b_3 \times (PropMD \times TotExp)$$

- interpret the F-statistic, R-squared value, standard error, and p-values.
- Evaluate the interaction term (PropMD * TotExp). What does this interaction tell us about the relationship between the number of MDs, healthcare spending, and life expectancy?

Discussion

- i How does the presence of more MDs amplify or diminish the effect of healthcare expenditures on life expectancy?
- ii What policy recommendations can be drawn from this analysis?

Question 5

Forecasting Life Expectancy with Interaction Terms

Task Using the multiple regression model from Question 4, forecast the life expectancy for a country where: The proportion of MDs is 0.03 (PropMD = 0.03). The total healthcare expenditure is 14 (TotExp = 14).

Discussion

- i Does this forecast seem realistic?
- ii Why or why not?
- iii Consider both the potential strengths and limitations of using this model for forecasting in real-world policy settings.

Problem 3

Question 1

Inventory Cost

Scenario A retail company is planning its inventory strategy for the upcoming year. They expect to sell 110 units of a high-demand product. The storage cost is \$3.75 per unit per year, and there is a fixed ordering cost of \$8.25 per order. The company wants to minimize its total inventory cost.

Task Using calculus, determine the optimal lot size (the number of units to order each time) and the number of orders the company should place per year to minimize total inventory costs. Assume that the total cost function is given by:

$$C(Q) = \frac{D}{Q} \times S + \frac{Q}{2} \times H$$

Where:

- D is total demand (110 units).
- Q is the order quantity.
- S is the fixed ordering cost per order (\$8.25).
- H is the holding cost per unit per year (\$3.75).

Question 2

Revenue Maximization

Scenario A company is running an online advertising campaign. The effectiveness of the campaign, in terms of revenue generated per day, is modeled by the function: $R(t) = -3150t^{-4} - 220t + 6530$ Where:

- $R(t)$ represents the revenue in dollars after t days of the campaign.

Task Determine the time t at which the revenue is maximized by finding the critical points of the revenue function and determining which point provides the maximum value. What is the maximum revenue the company can expect from this campaign?

Question 3

Demand Area Under Curve

Scenario A company sells a product at a price that decreases over time according to the linear demand function: Where: $P(x) = 2x - 9.3$

$P(x)$ is the price in dollars, and x is the quantity sold.

Task The company is interested in calculating the total revenue generated by this product between two quantity levels, $x_1 = 2$ and $x_2 = 5$, where the price still generates sales. Compute the area under the demand curve between these two points, representing the total revenue generated over this range.

Question 4

Profit Optimization

Scenario A beauty supply store sells flat irons, and the profit function associated with selling x flat irons is given by:

$$\Pi(x) = x \ln(9x) - \frac{x^6}{6}$$

Where:

- $\Pi(x)$ is the profit in dollars.

Task Use calculus to find the value of x that maximizes profit. Calculate the maximum profit that can be achieved and determine if this optimal sales level is feasible given market conditions.

Question 5

Spending Behavior

Scenario A market research firm is analyzing the spending behavior of customers in a retail store. The spending behavior is modeled by the probability density function: $f(x) = \frac{1}{6x}$

Where x represents spending in dollars.

Task Determine whether this function is a valid probability density function over the interval $[1, e^6]$. If it is, calculate the probability that a customer spends between \$1 and e^6 .

Question 6

Market Share Estimation

Scenario An electronics company is analyzing its market share over a certain period. The rate of market penetration is given by: $\frac{dN}{dt} = \frac{500}{t^4+10}$

Where $N(t)$ is the cumulative market share at time t .

Task Integrate this function to find the cumulative market share $N(t)$ after t days, given that the initial market share $N(1) = 6530$. What will the market share be after 10 days?

Problem 4

Business Optimization

As a data scientist at a consultancy firm, you are tasked with optimizing various business functions to improve efficiency and profitability. Taylor Series expansions are a powerful tool to approximate complex functions, allowing for simpler calculations and more straightforward decision-making. This week, you will work on Taylor Series expansions of popular functions commonly encountered in business scenarios.

Question 1

Revenue and Cost

Scenario A company's revenue from a product can be approximated by the function $R(x) = e^x$, where x is the number of units sold. The cost of production is given by $C(x) = \ln(1+x)$. The company wants to maximize its profit, defined as $\Pi(x) = R(x) - C(x)$.

Task

1. **Approximate the Revenue Function:** Use the Taylor Series expansion around $x = 0$ (Maclaurin series) to approximate the revenue function $R(x) = e^x$ up to the second degree. Explain why this approximation might be useful in a business context.
2. **Approximate the Cost Function:** Similarly, approximate the cost function $C(x) = \ln(1+x)$ using its Maclaurin series expansion up to the second degree. Discuss the implications of this approximation for decision-making in production.
3. **Linear vs. Nonlinear Optimization:** Using the Taylor Series expansions, approximate the profit function $\Pi(x)$. Compare the optimization results when using the linear approximations versus the original nonlinear functions. What are the differences, and when might it be more appropriate to use the approximation?

Submission Provide your solutions using R-Markdown. Include the Taylor Series expansions, the approximated functions, and a discussion of the implications of using these approximations for business decision-making.

Question 2

Financial Modeling

Scenario A financial analyst is modeling the risk associated with a new investment. The risk is proportional to the square root of the invested amount, modeled as $f(x) = \sqrt{x}$, where x is the amount invested. However, to simplify calculations, the analyst wants to use a Taylor Series expansion to approximate this function for small investments.

Task

1. **Maclaurin Series Expansion:** Derive the Taylor Series expansion of $f(x) = \sqrt{x}$ around $x = 0$ up to the second degree.
2. **Practical Application:** Use the derived series to approximate the risk for small investment amounts (e.g., when x is small). Compare the approximated risk with the actual function values for small and moderate investments. Discuss when this approximation might be useful in financial modeling.
3. **Optimization Scenario:** Suppose the goal is to minimize risk while maintaining a certain level of investment return. Using the Taylor Series approximation, suggest an optimal investment amount x that balances risk and return.

Submission Present your results in R-Markdown, including the Taylor Series expansions, comparisons between the original and approximated functions, and your recommendations based on the analysis.

Question 3

Inventory Management

Scenario In a manufacturing process, the demand for a product decreases as the price increases, modeled by $D(p) = 1 - p$, where p is the price. The cost associated with producing and selling the product is modeled as $C(p) = e^p$. The company wants to maximize its profit, which is the difference between revenue and cost.

Task

1. **Taylor Series Expansion:** Expand the cost function $C(p) = e^p$ into a Taylor Series around $p = 0$ up to the second degree. Discuss why approximating the cost function might be useful in a pricing strategy.
2. **Approximating Profit:** Using the Taylor Series expansion, approximate the profit function $\Pi(p) = pD(p) - C(p)$. Compare the results when using the original nonlinear cost function versus the approximated cost function. What differences do you observe, and when might the approximation be sufficient?
3. **Pricing Strategy:** Based on the Taylor Series approximation, suggest a pricing strategy that could maximize profit. Explain how the Taylor Series approximation helps in making this decision.

Submission Include your analysis in R-Markdown, with Taylor Series expansions, comparisons of the approximated and original functions, and a discussion of the implications for pricing strategy.

Question 4

Economic Forecasting

Scenario An economist is forecasting economic growth, which can be modeled by the logarithmic function $G(x) = \ln(1 + x)$, where x represents investment in infrastructure. The government wants to predict growth under different levels of investment.

Task

i. Maclaurin Series Expansion: Derive the Maclaurin Series expansion of $G(x) = \ln(x + 1)$ up to the second degree. Explain the significance of using this approximation for small values of x_x in economic forecasting.

ii. Approximation of Growth: Use the Taylor Series to approximate the growth for small investments. Compare this approximation with the actual growth function. Discuss the accuracy of the approximation for different ranges of x_x .

iii. Policy Recommendation: Using the approximation, recommend a level of investment that could achieve a target growth rate. Discuss the limitations of using Taylor Series approximations for such policy recommendations.

Submission Provide your answers in R-Markdown, with the Taylor Series expansions, comparisons between the approximated and original functions, and your investment recommendations.

Problem 5

Profit, Cost, & Pricing

Question 1

Profit Maximization

Scenario A company produces two products, A and B. The profit function for the two products is given by $\Pi(x, y) = 30x - 2x^2 - 3xy + 24y - 4y^2$

Where:

- x is the quantity of Product A produced and sold.
- y is the quantity of Product B produced and sold.
- $\Pi(x, y)$ is the profit in dollars.

Task Find all local maxima, local minima, and saddle points for the profit function $\Pi(x, y)$. Write your answer(s) in the form $(x, y, \Pi(x, y))$. Separate multiple points with a comma.

Discussion Discuss the implications of the results for the company's production strategy. Which production levels maximize profit, and what risks are associated with the saddle points?

Question 2

Pricing Strategy

Scenario A supermarket sells two competing brands of a product: Brand X and Brand Y. The store manager estimates that the demand for these brands depends on their prices, given by the functions:

- Demand for Brand X: $D_x(x, y) = 120 - 15x + 10y$
- Demand for Brand Y: $D_y(x, y) = 80 + 5x - 20y$

Where:

- x is the price of Brand X in dollars.
- y is the price of Brand Y in dollars.
- $D_x(x, y)$ and $D_y(x, y)$ are the quantities demanded for Brand X and Brand Y, respectively.

Task

1. **Revenue Function:** Find the revenue function $R(x,y)$ for both brands combined.
2. **Optimal Pricing:** Determine the prices x and y that maximize the store's total revenue. Are there any saddle points to consider in the pricing strategy?

Discussion Explain the significance of the optimal pricing strategy and how it can be applied in a competitive retail environment.

Question 3

Cost Minimization

Scenario A manufacturing company operates two plants, one in New York and one in Chicago. The company needs to produce a total of 200 units of a product each week. The total weekly cost of production is given by $C(x, y) = \frac{1}{8}x^2 + \frac{1}{10}y^2 + 12x + 18y + 1500$

Where:

- x is the number of units produced in New York.
- y is the number of units produced in Chicago.
- $C(x, y)$ is the total cost in dollars.

Task

1. Determine how many units should be produced in each plant to minimize the total weekly cost.
2. What is the minimized total cost, and how does the distribution of production between the two plants affect overall efficiency?

Discussion Discuss the benefits of this cost-minimization strategy and any practical considerations that might influence the allocation of production between the two plants.

Question 4

Marketing Mix

Scenario A company is launching a marketing campaign that involves spending on online ads (x) and television ads (y). The effectiveness of the campaign, measured in customer reach, is modeled by the function $E(x, y) = 500x + 700y - 5x^2 - 10xy - 8y^2$

Where:

- x is the amount spent on online ads (in thousands of dollars).
- y is the amount spent on television ads (in thousands of dollars).
- $E(x, y)$ is the estimated customer reach.

Task

1. Find the spending levels for online and television ads that maximize customer reach.
2. Identify any saddle points and discuss how they could affect the marketing strategy.

Discussion Explain how the results can be used to allocate the marketing budget effectively and what the company should consider if it encounters saddle points in the optimization.

As always, include a professional pdf only with all code, text, graphics, explanations, etc.