STAT632 Project (Advertisement Sales Dataset)

David Teng

2025-04-30

About Dataset

Advertisement Sales Dataset

The Advertisement Sales dataset is a collection of data points used to analyze the impact of advertising on sales. This dataset consists of 200 entries, each representing a unique observation with data on various types of media advertising and corresponding sales figures.

Load Libraries and Data

```r
# Load necessary libraries
library(MASS)          # for boxcox
library(glmnet)        # for LASSO
```

```
Loading required package: Matrix
```

```
Loaded glmnet 4.1-8
```

```r
library(randomForest)  # for Random Forest
```

```
randomForest 4.7-1.2
```

```
Type rfNews() to see new features/changes/bug fixes.
```

```r
library(car)            # for VIF
```

```
Loading required package: carData
```

```r
library(ggplot2)        # for nice plots
```

```

Attaching package: 'ggplot2'

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

    margin
```

```r
library(caret)          # for model validation
```

```
Loading required package: lattice
```

```r
library(dplyr)          # for data manipulation
```

```

Attaching package: 'dplyr'

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

    recode

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

    combine

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

    select

The following objects are masked from 'package:stats':

    filter, lag
```

```
The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

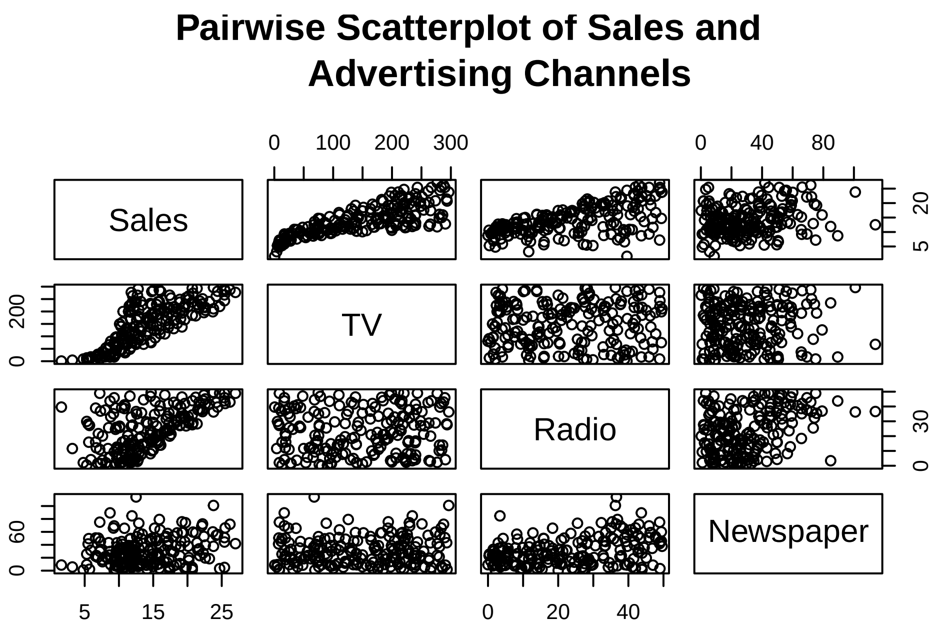
```r
# Load data
adver <- read.csv("Advertising And Sales.csv")

# Quick overview
summary(adver)
```

```
       ID               TV             Radio          Newspaper     
 Min.   :  1.00   Min.   :  0.70   Min.   : 0.00   Min.   :  0.30  
 1st Qu.: 50.75   1st Qu.: 74.38   1st Qu.:10.07   1st Qu.: 12.75  
 Median :100.50   Median :149.75   Median :22.90   Median : 25.75  
 Mean   :100.50   Mean   :147.03   Mean   :23.29   Mean   : 30.55  
 3rd Qu.:150.25   3rd Qu.:218.82   3rd Qu.:36.52   3rd Qu.: 45.10  
 Max.   :200.00   Max.   :296.40   Max.   :49.60   Max.   :114.00  
     Sales      
 Min.   : 1.60  
 1st Qu.:10.40  
 Median :12.90  
 Mean   :14.04  
 3rd Qu.:17.40  
 Max.   :27.00  
```

```r
str(adver)
```

```
'data.frame':   200 obs. of  5 variables:
 $ ID       : int  1 2 3 4 5 6 7 8 9 10 ...
 $ TV       : num  230.1 44.5 17.2 151.5 180.8 ...
 $ Radio    : num  37.8 39.3 45.9 41.3 12.8 48.9 32.8 19.6 2.1 2.6 ...
 $ Newspaper: num  69.2 45.1 69.3 58.5 58.4 75 23.5 11.6 1 21.2 ...
 $ Sales    : num  22.1 10.4 9.3 18.5 12.9 7.2 11.8 13.2 4.8 10.6 ...
```

```r
# Pairwise scatterplot with title
pairs(Sales ~ TV + Radio + Newspaper, data = adver,
      main = "Pairwise Scatterplot of Sales and
      Advertising Channels")
```



Data Preparation

```r
# Check for missing values
colSums(is.na(adver))
```

```
       ID        TV     Radio Newspaper     Sales
        0         0         0         0         0
```

Base Multiple Linear Regression

```r
# Base Multiple Linear Regression
lm1 <- lm(Sales ~ TV + Radio + Newspaper, data = adver)
summary(lm1)
```

```
Call:
lm(formula = Sales ~ TV + Radio + Newspaper, data = adver)

Residuals:
    Min      1Q  Median      3Q     Max
```

```
-8.8335 -0.8662  0.2411  1.1927  3.4411

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  3.003556   0.313296   9.587   <2e-16 ***
TV           0.045686   0.001402  32.583   <2e-16 ***
Radio        0.187110   0.008649  21.634   <2e-16 ***
Newspaper   -0.001330   0.005905  -0.225    0.822    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.695 on 196 degrees of freedom
Multiple R-squared:  0.8958,	Adjusted R-squared:  0.8942 
F-statistic: 561.4 on 3 and 196 DF,  p-value: < 2.2e-16
```

Interpretation of Base Multiple Linear Regression:

We fitted a multiple linear regression model to predict **Sales** based on **TV**, **Radio**, and **Newspaper** advertising budgets.

- The **Intercept** is estimated at **3.00** ($p < 0.001$), meaning that when advertising budgets are zero, the expected sales would be about 3,000 units.
- The **TV advertising budget** has a **positive and significant** effect on Sales.

  Each additional thousand dollars spent on TV is associated with an **increase of about 45.7 units** in Sales, **holding other factors constant** ($p < 0.001$).
- The **Radio advertising budget** also has a **positive and significant** effect on Sales.

  Each additional thousand dollars spent on Radio is associated with an **increase of about 187.1 units** in Sales ($p < 0.001$).
- The **Newspaper advertising budget** is **not statistically significant** ($p = 0.822$), suggesting that spending on Newspaper ads does **not have a meaningful effect** on Sales in this model.

Goodness of Fit:

- The **Multiple R-squared** is **0.8958**, meaning the model explains about **89.6%** of the variance in Sales.
- The **Adjusted R-squared** is **0.8942**, which adjusts for the number of predictors and confirms the model still fits the data very well.

- The overall **F-statistic** is highly significant (p < 2.2e-16), indicating that the model provides a better fit than a model with no predictors.

Conclusion:

The model shows that TV and Radio advertising significantly increase sales, while Newspaper advertising does not. The model explains 89.6% of the variance in sales, and overall, it fits the data very well (F-test p < 0.001).

Reduced Model (remove Newspaper)

```r
# Reduced Model (remove Newspaper)
lm2 <- lm(Sales ~ TV + Radio, data = adver)
summary(lm2)
```

```
Call:
lm(formula = Sales ~ TV + Radio, data = adver)

Residuals:
    Min      1Q  Median      3Q     Max 
-8.7951 -0.8621  0.2422  1.1749  3.4344 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 2.980757   0.295772   10.08   <2e-16 ***
TV          0.045674   0.001398   32.68   <2e-16 ***
Radio       0.186423   0.008073   23.09   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.691 on 197 degrees of freedom
Multiple R-squared:  0.8957,	Adjusted R-squared:  0.8947 
F-statistic: 846.2 on 2 and 197 DF,  p-value: < 2.2e-16
```

Interpretation of Reduced Model (TV + Radio only):

We fitted a reduced multiple linear regression model to predict **Sales** using only **TV** and **Radio** advertising budgets (after removing Newspaper).

- The **Intercept** is estimated at **2.98** ($p < 0.001$), meaning that when TV and Radio advertising expenditures are zero, the expected sales would be about **2,980 units**.
- The **TV advertising budget** remains a **positive and highly significant predictor** of Sales.

  Each additional thousand dollars spent on TV advertising is associated with an **increase of approximately 45.7 units** in Sales, **holding Radio constant** ($p < 0.001$).
- The **Radio advertising budget** also remains **positive and highly significant**.

  Each additional thousand dollars spent on Radio advertising is associated with an **increase of approximately 186.4 units** in Sales ($p < 0.001$).

Goodness of Fit:

- The **Multiple R-squared** is **0.8957**, indicating that about **89.6%** of the variance in Sales is explained by TV and Radio budgets.
- The **Adjusted R-squared** is **0.8947**, very close to the full model, suggesting that removing Newspaper **did not harm model fit**.
- The model's **F-statistic** is **highly significant** ($p < 2.2e\text{-}16$), showing the model overall is statistically significant.

Model Comparison:

- Compared to the full model (TV + Radio + Newspaper), the reduced model achieves **almost identical R-squared** with fewer predictors.
- Based on the **partial F-test** and **adjusted R-squared**, we conclude that **Newspaper** advertising is **not necessary** for predicting Sales.

Conclusion:

The reduced model including only TV and Radio advertising performs just as well as the full model. Both TV and Radio advertising expenditures have significant positive effects on Sales, while Newspaper advertising was found to be unnecessary. The reduced model explains about 89.6% of the variance in Sales and provides a simpler, equally effective prediction model.

Compare Models: Full vs Reduced:

Hypotheses for Model Comparison:

$H_0 : \beta_{Newspaper} = 0$ (The coefficient for **Newspaper** is equal to zero which means Newspaper does **not** improve the model.)

vs. $H_1 : \beta_{Newspaper} \neq 0$

(The coefficient for **Newspaper** is **not** equal to zero which means Newspaper **does** improve the model.)

```r
# Compare Models: Full vs Reduced
anova(lm2, lm1) # partial F-test
```

```
Analysis of Variance Table

Model 1: Sales ~ TV + Radio
Model 2: Sales ~ TV + Radio + Newspaper
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    197 563.09
2    196 562.95  1   0.14567 0.0507 0.8221
```

Interpretation of Model Comparison (Full vs Reduced):

We conducted a **partial F-test** to formally compare the full model (**Sales** ~ **TV** + **Radio** + **Newspaper**) with the reduced model (**Sales** ~ **TV** + **Radio**).

From the ANOVA table:

- The test statistic is **F = 0.0507**, with a corresponding **p-value = 0.8221**.
- The p-value is **much greater than 0.05**, meaning we **fail to reject** the null hypothesis.

Interpretation:

- There is **no significant evidence** that adding **Newspaper** as a predictor improves the model.
- Therefore, the **simpler model with only TV and Radio** is preferred.

Additional Note:

- The Residual Sum of Squares (RSS) only **slightly decreased** from **563.09** to **562.95** after adding Newspaper, which is not meaningful.
- This further confirms that **Newspaper** is not a useful predictor for Sales.

Conclusion:

Since the partial F-test (p = 0.8221), we **fail to reject** H_0.

This means **Newspaper does not significantly improve** the model. Thus, the reduced model with only TV and Radio is sufficient.

Adjusted R-squared comparison

```r
# Adjusted R-squared comparison
summary(lm1)$adj.r.squared # Full model (TV + Radio + Newspaper
```

```
[1] 0.8941635
```

```r
summary(lm2)$adj.r.squared # Reduced model (TV + Radio only)
```

```
[1] 0.8946735
```

Adjusted R-squared Comparison Interpretation:

The **adjusted R-squared** for the reduced model (**0.8947**) is **slightly higher** than that of the full model (**0.8942**).

- **Adjusted R-squared** adjusts for the number of predictors in the model.
- A **higher adjusted R-squared** suggests that the reduced model **fits the data better**, even though it uses **fewer predictors**.
- Therefore, the model including only **TV** and **Radio** provides a **better and simpler fit** than the model that also includes **Newspaper**.

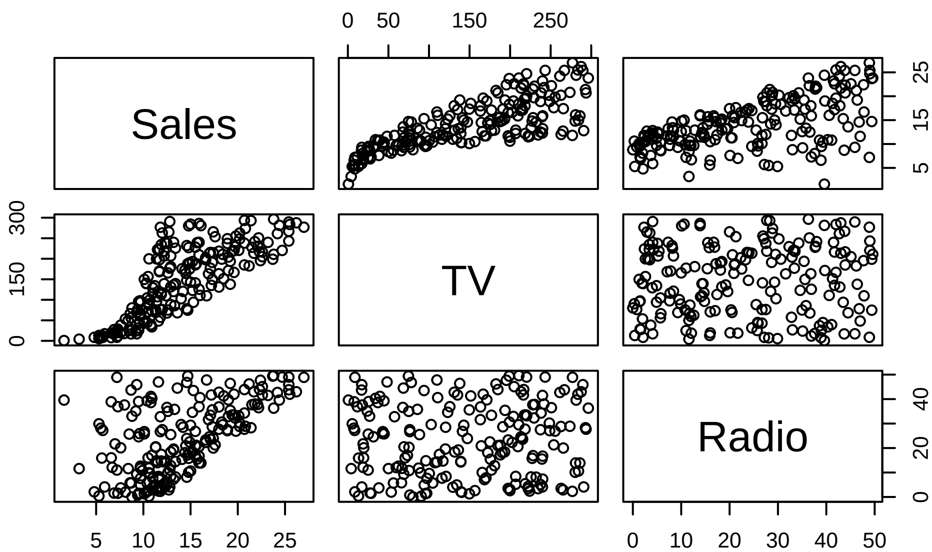
Conclusion:

The reduced model (TV + Radio) has a slightly higher adjusted R-squared than the full model, indicating a better fit with fewer predictors.

Result: Slightly better adjusted R^2 for reduced model → remove Newspaper

Pairwise scatterplot

```
pairs(Sales ~ TV + Radio, data = adver)
```



Interpretation of Pairwise Scatterplot (Sales, TV, Radio):

The pairwise scatterplot shows the relationships between **Sales**, **TV**, and **Radio** advertising:

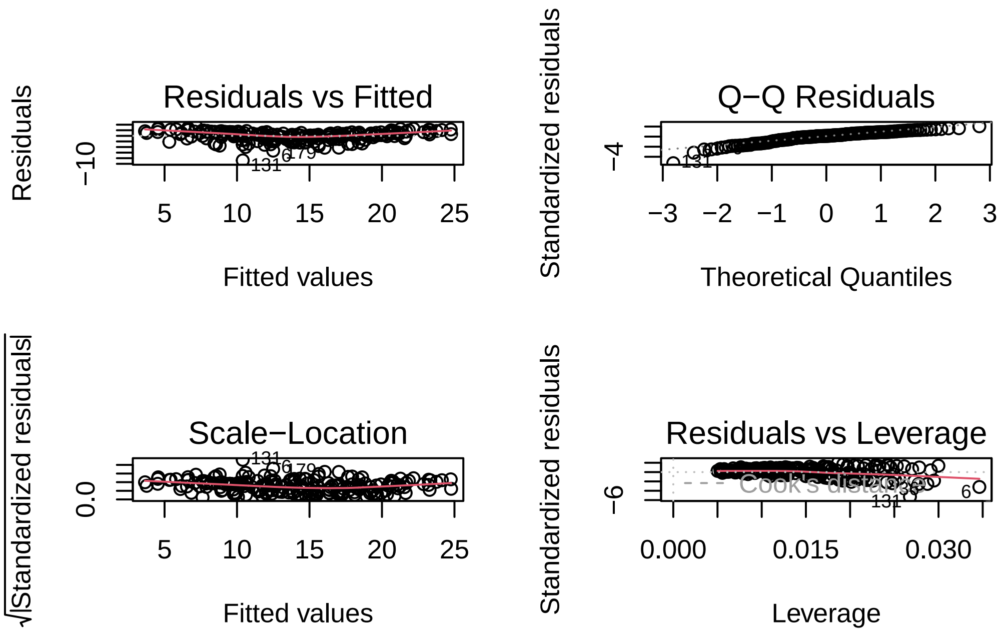
- **Sales vs TV**: There is a **strong positive linear relationship**. As spending on TV advertising increases, Sales also tend to increase. The pattern is clear and linear, supporting the use of TV as a predictor in a linear regression model.
- **Sales vs Radio**: A **moderate positive linear relationship** is also observed. Though more spread out than the TV relationship, the trend is still upward, suggesting Radio advertising has a meaningful impact on Sales.
- **TV vs Radio**: The scatterplot shows **no strong correlation** between TV and Radio advertising budgets. The points are scattered without a clear pattern, suggesting that TV and Radio are **not highly collinear**, which is good for regression modeling.

Conclusion:

Sales shows strong positive correlation with TV advertising and moderate positive correlation with Radio advertising. TV and Radio budgets appear to be largely independent.

Diagnostic Plots for lm2

```
# Diagnostic Plots for lm2
par(mfrow=c(2,2))
plot(lm2)
```



Diagnostic Plots Interpretation (for Reduced Model):

These diagnostic plots help assess the assumptions of the multiple linear regression model:

1. Residuals vs Fitted

- This plot checks for **linearity** and **homoscedasticity**.

- The residuals appear to be randomly scattered around the horizontal line, indicating that:

  The relationship between predictors and response is likely **linear**.

  There is **no clear pattern**, suggesting **constant variance** (no heteroscedasticity).

2. Normal Q-Q Plot

- This plot checks for **normality of residuals**.
- The residual points mostly fall along the straight line, indicating that the residuals are **approximately normally distributed**.

3. Scale-Location Plot

- This plot also checks for **homoscedasticity**, using standardized residuals.
- The red line is mostly flat and the spread of residuals is consistent across fitted values, suggesting **homogeneity of variance**.

4. Residuals vs Leverage

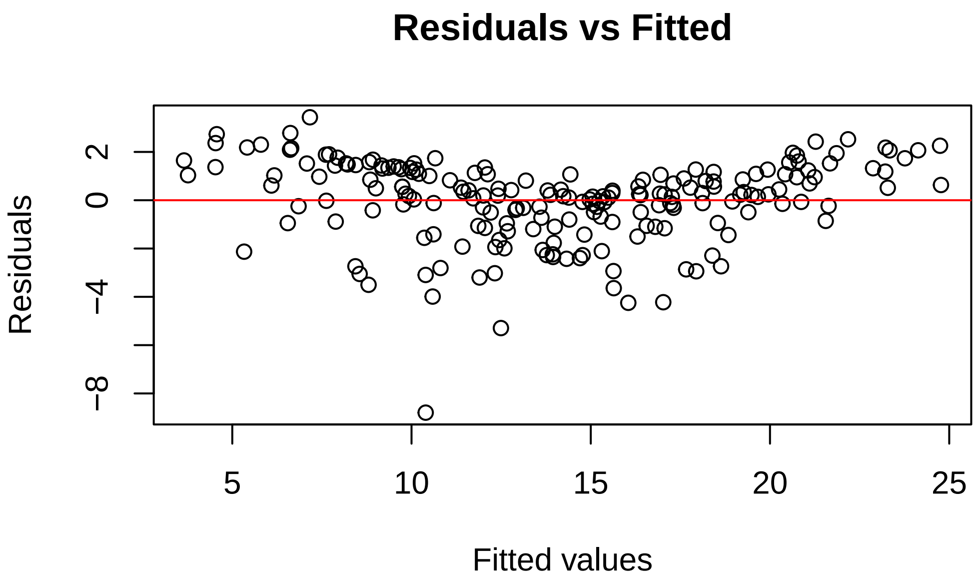
- This plot identifies **influential points** that may disproportionately affect the model.
- There are no points with **unusually high leverage** or **extreme residuals**, indicating that there are **no strong outliers or influential observations**.

Conclusion:

The diagnostic plots suggest that the reduced model meets the assumptions of linearity, normality, constant variance, and no influential outliers. Thus, the model appears appropriate for inference and prediction.

Residuals vs Fitted for lm2

```
# Residuals vs Fitted for lm2
plot(fitted(lm2), resid(lm2),
     xlab="Fitted values", ylab="Residuals",
     main="Residuals vs Fitted")
abline(h=0, col="red")
```

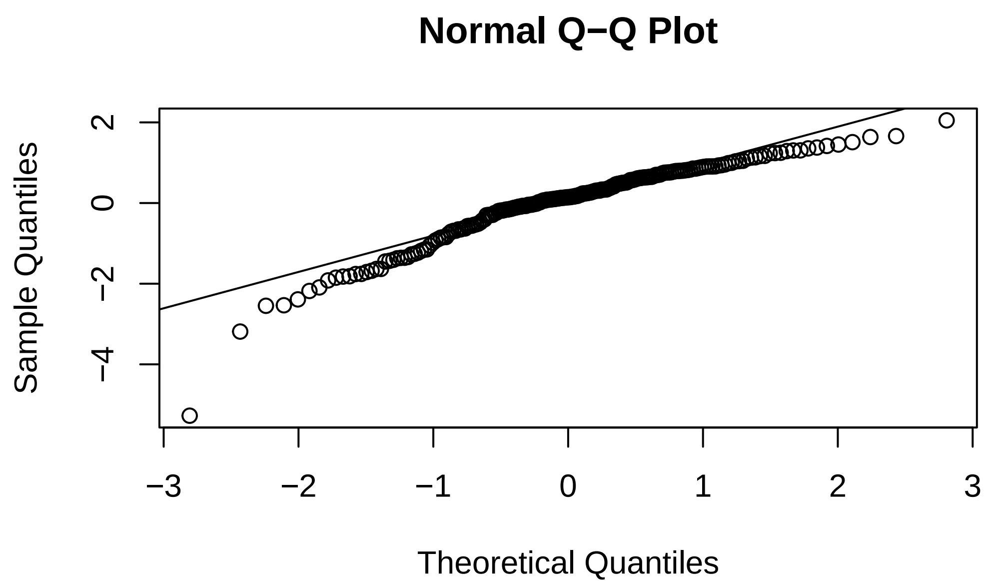


Interpretation:

The Residuals vs Fitted plot shows that residuals are randomly scattered around zero with no strong pattern, supporting the assumptions of linearity and constant variance.

QQ plot for residuals (Normality)

```
# QQ plot for residuals (Normality)
qqnorm(rstandard(lm2))
qqline(rstandard(lm2))
```

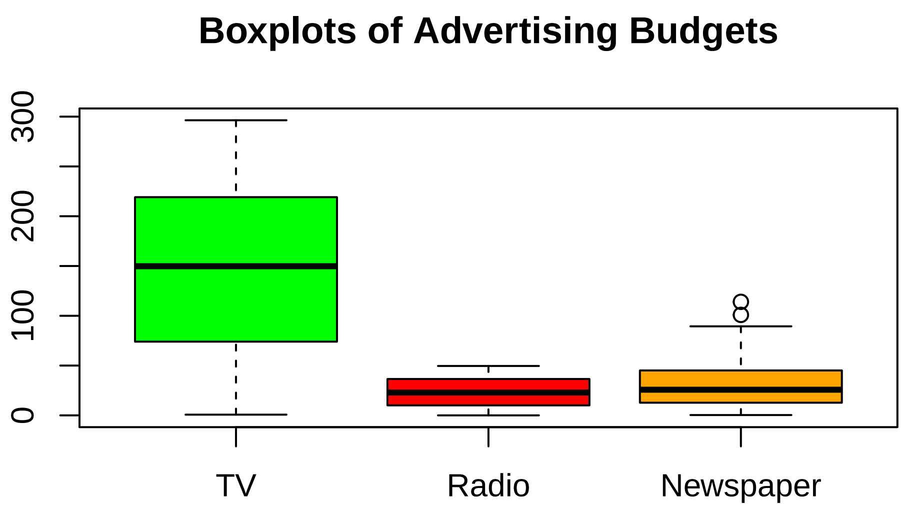


Interpretation:

The Q-Q plot suggests that the residuals are **approximately normally distributed**, with **minor deviations at the tails**. This does **not seriously violate** the normality assumption required for multiple linear regression.

Boxplots to check for outliers in predictors

```
# Boxplots to check for outliers in predictors
boxplot(adver[, c("TV", "Radio", "Newspaper")],
        main="Boxplots of Advertising Budgets",
        col=c("green", "red", "orange"))
```



Interpretation:

The boxplots show that **TV has the largest budget range**, while **Newspaper advertising contains a few outliers**. No extreme values are observed for TV or Radio. This insight helps explain why Newspaper may not be a strong predictor in the regression model — its distribution is more scattered and includes outlying values.

Why Use a Box-Cox Transformation for `lm1`?

For Model (`lm1: Sales ~ TV + Radio + Newspaper`)

- We already checked the **residuals vs fitted plot** and **QQ plot**, which were **mostly okay**, but:

  There was **some non-linearity** and **slight skewness** in the residuals.

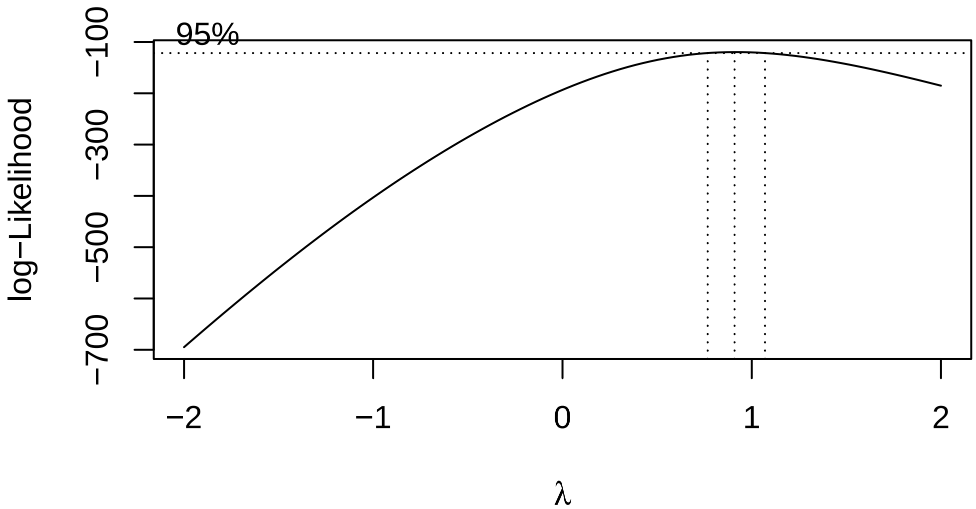
- So using `boxcox()` helps :

  Confirm whether transformation is needed, and

  Find the best power transformation (e.g., log, sqrt, etc.) to improve the model.

```
# Box-Cox transformation for lm1
boxcox(lm1, lambda = seq(-2, 2, 0.1),
       main = "Box-Cox Transformation for Sales")
```

```
Warning: In lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...) :
 extra argument 'main' will be disregarded
```



Interpretation:

The Box-Cox transformation plot indicates that the optimal λ is close to 1, and **no transformation of the response variable is necessary**. This supports using the original Sales variable in the multiple linear regression model.

Check Influential Observations

```
# Check Influential Observations
influence.measures(lm2)
```

```
Influence measures of
	 lm(formula = Sales ~ TV + Radio, data = adver) :

      dfb.1_    dfb.TV  dfb.Radi    dffit cov.r   cook.d     hat inf
1   -5.43e-02  6.07e-02  6.13e-02  0.11080 1.016 4.09e-03 0.01398
2   -4.54e-02  1.04e-01 -9.51e-02 -0.16031 1.014 8.55e-03 0.01877
3   -6.47e-02  2.10e-01 -2.11e-01 -0.31829 0.995 3.34e-02 0.02949
4   -1.35e-02 -6.43e-04  4.62e-02  0.05987 1.024 1.20e-03 0.01237
5   -1.73e-02 -1.33e-02  2.23e-02 -0.03966 1.021 5.26e-04 0.00845
6   -1.03e-01  4.03e-01 -4.28e-01 -0.61818 0.897 1.21e-01 0.03462   *
7    2.35e-03 -3.59e-03  2.32e-03  0.00532 1.029 9.49e-06 0.01292
8    3.43e-02 -1.35e-02 -1.04e-02  0.04847 1.015 7.86e-04 0.00576
9    1.03e-01 -6.83e-02 -5.95e-02  0.10326 1.037 3.57e-03 0.02702
10  -7.05e-02 -5.91e-02  1.21e-01 -0.15710 1.011 8.21e-03 0.01715
```

11	1.12e-01	-5.67e-02	-7.27e-02	0.11475	1.019	4.39e-03	0.01582	
12	-8.50e-04	4.61e-03	1.76e-05	0.00745	1.024	1.86e-05	0.00812	
13	-5.28e-02	8.94e-02	-5.29e-02	-0.11788	1.024	4.64e-03	0.01921	
14	5.09e-02	-1.86e-02	-3.68e-02	0.05545	1.024	1.03e-03	0.01194	
15	-1.00e-02	1.50e-02	1.45e-02	0.03201	1.023	3.43e-04	0.00909	
16	-6.50e-02	3.22e-02	1.10e-01	0.13550	1.021	6.12e-03	0.01965	
17	-8.11e-03	1.66e-02	-1.61e-02	-0.02815	1.029	2.65e-04	0.01380	
18	-6.51e-02	7.63e-02	5.12e-02	0.10716	1.031	3.84e-03	0.02242	
19	6.29e-02	-5.07e-02	-7.70e-03	0.07657	1.015	1.96e-03	0.00922	
20	6.90e-03	1.43e-05	7.50e-04	0.01830	1.019	1.12e-04	0.00501	
21	1.63e-03	-4.12e-03	-1.26e-03	-0.00669	1.024	1.50e-05	0.00878	
22	-4.28e-02	-1.09e-01	1.25e-01	-0.18873	1.006	1.18e-02	0.01886	
23	-6.98e-02	6.19e-02	1.65e-02	-0.07655	1.029	1.96e-03	0.01804	
24	-1.59e-03	-4.36e-02	2.17e-02	-0.06530	1.020	1.43e-03	0.01068	
25	9.60e-02	-6.13e-02	-4.29e-02	0.10041	1.015	3.36e-03	0.01211	
26	-4.86e-02	-2.25e-01	2.23e-01	-0.34616	0.967	3.92e-02	0.02413	
27	2.46e-04	-9.06e-05	5.19e-04	0.00137	1.021	6.31e-07	0.00585	
28	2.32e-03	-5.46e-02	2.49e-02	-0.07659	1.020	1.96e-03	0.01219	
29	1.12e-02	-2.46e-02	-3.98e-03	-0.03281	1.027	3.60e-04	0.01224	
30	7.12e-02	-4.79e-02	-2.44e-02	0.07812	1.016	2.04e-03	0.00995	
31	9.19e-03	-1.68e-02	-2.41e-03	-0.01985	1.036	1.32e-04	0.01979	
32	1.97e-02	-8.20e-03	-8.17e-03	0.02484	1.021	2.07e-04	0.00650	
33	1.35e-01	-4.04e-02	-1.16e-01	0.14934	1.013	7.42e-03	0.01703	
34	2.30e-02	-8.55e-02	1.84e-02	-0.10596	1.019	3.75e-03	0.01503	
35	1.35e-01	-4.15e-02	-1.16e-01	0.14916	1.013	7.41e-03	0.01722	
36	-1.30e-02	-3.23e-01	2.57e-01	-0.44209	0.946	6.33e-02	0.02874	*
37	-1.19e-01	1.17e-01	1.16e-01	0.19105	1.016	1.21e-02	0.02331	
38	3.47e-03	3.65e-02	-6.99e-02	-0.08629	1.037	2.49e-03	0.02497	
39	7.81e-03	-8.99e-03	2.19e-03	0.01172	1.028	4.60e-05	0.01281	
40	-3.65e-02	4.07e-02	4.19e-02	0.07549	1.023	1.91e-03	0.01369	
41	1.51e-04	5.84e-03	-9.28e-04	0.01071	1.022	3.84e-05	0.00715	
42	1.88e-03	-2.51e-03	-5.34e-03	-0.01007	1.023	3.39e-05	0.00781	
43	3.28e-02	-6.18e-02	-7.30e-03	-0.07249	1.032	1.76e-03	0.01984	
44	-2.52e-02	-3.51e-02	4.85e-02	-0.07454	1.022	1.86e-03	0.01290	
45	-2.21e-02	2.54e-02	-4.32e-03	-0.03113	1.030	3.25e-04	0.01543	
46	-2.52e-03	-3.79e-03	8.24e-04	-0.01207	1.021	4.88e-05	0.00556	
47	9.82e-02	-4.39e-02	-6.14e-02	0.10512	1.011	3.68e-03	0.01099	
48	-6.78e-02	6.60e-02	7.58e-02	0.12214	1.021	4.98e-03	0.01770	
49	-5.61e-03	-6.14e-02	3.55e-02	-0.09399	1.014	2.95e-03	0.01094	
50	9.33e-02	-5.60e-02	-4.57e-02	0.09731	1.016	3.16e-03	0.01203	
51	-4.42e-02	-3.79e-02	7.63e-02	-0.09965	1.023	3.32e-03	0.01667	
52	7.46e-02	-2.80e-02	-5.07e-02	0.08213	1.016	2.25e-03	0.01047	
53	-7.27e-02	6.18e-02	9.98e-02	0.14659	1.010	7.15e-03	0.01544	

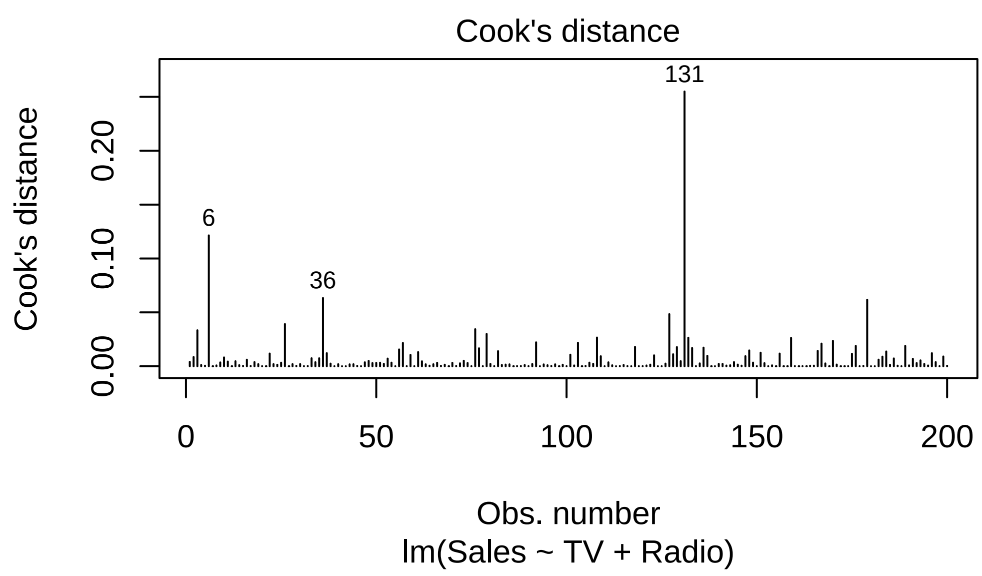
54	-4.27e-02	1.77e-02	8.20e-02	0.10065	1.024	3.38e-03	0.01747
55	4.45e-03	-8.36e-03	-1.85e-03	-0.01071	1.030	3.84e-05	0.01455
56	-1.09e-01	5.26e-02	1.80e-01	0.21702	1.005	1.56e-02	0.02177
57	-1.68e-01	2.16e-01	-5.47e-02	-0.25643	0.984	2.17e-02	0.01917
58	1.15e-02	-1.95e-03	-4.73e-03	0.01834	1.020	1.13e-04	0.00544
59	-9.55e-02	5.37e-02	1.44e-01	0.17793	1.018	1.05e-02	0.02281
60	-4.19e-03	8.98e-03	4.70e-03	0.01623	1.024	8.82e-05	0.00847
61	1.95e-01	-9.97e-02	-1.36e-01	0.19915	1.007	1.32e-02	0.02042
62	-7.11e-02	7.13e-02	6.98e-02	0.11733	1.028	4.60e-03	0.02152
63	-3.36e-04	-5.16e-02	2.73e-02	-0.07364	1.022	1.81e-03	0.01252
64	1.68e-02	-1.85e-02	1.55e-02	0.04140	1.019	5.73e-04	0.00738
65	-1.15e-02	-1.16e-02	5.92e-02	0.07457	1.024	1.86e-03	0.01399
66	9.52e-02	-5.21e-02	-5.42e-02	0.09843	1.018	3.23e-03	0.01313
67	2.60e-02	-2.84e-02	3.46e-03	0.03531	1.029	4.18e-04	0.01423
68	4.66e-02	-3.23e-03	-3.35e-02	0.06628	1.012	1.47e-03	0.00677
69	9.80e-04	-2.21e-03	-4.76e-04	-0.00312	1.026	3.27e-06	0.01081
70	-5.05e-02	3.86e-02	7.05e-02	0.09771	1.025	3.19e-03	0.01737
71	-6.38e-03	1.27e-02	1.01e-02	0.02755	1.022	2.54e-04	0.00790
72	7.71e-02	-2.94e-02	-4.27e-02	0.09058	1.007	2.73e-03	0.00764
73	-6.22e-02	9.56e-02	-4.87e-02	-0.12413	1.020	5.14e-03	0.01755
74	7.60e-02	-7.63e-03	-7.46e-02	0.09505	1.019	3.02e-03	0.01354
75	2.11e-03	-1.02e-02	-5.86e-04	-0.01669	1.023	9.33e-05	0.00801
76	-8.10e-02	2.22e-01	-2.03e-01	-0.32344	0.986	3.44e-02	0.02726
77	2.23e-01	-1.34e-01	-1.41e-01	0.22420	1.009	1.67e-02	0.02433
78	6.74e-03	-5.58e-03	6.24e-03	0.01876	1.021	1.18e-04	0.00616
79	-1.85e-01	2.53e-01	-8.15e-02	-0.30247	0.969	3.00e-02	0.02013
80	6.90e-02	-1.65e-02	-5.61e-02	0.08049	1.018	2.16e-03	0.01098
81	1.20e-02	-1.24e-02	4.11e-03	0.01965	1.024	1.29e-04	0.00878
82	-4.74e-02	-1.19e-01	1.39e-01	-0.20562	1.004	1.40e-02	0.02008
83	4.97e-02	-3.82e-02	-7.12e-03	0.06062	1.018	1.23e-03	0.00862
84	-5.32e-03	3.41e-02	-5.06e-02	-0.06862	1.033	1.58e-03	0.02023
85	-3.64e-02	2.83e-02	5.20e-02	0.07291	1.027	1.78e-03	0.01629
86	-3.62e-04	-8.25e-04	5.33e-04	-0.00176	1.023	1.04e-06	0.00710
87	1.34e-02	-1.45e-02	5.69e-03	0.02297	1.024	1.77e-04	0.00896
88	-2.15e-04	-8.19e-03	1.99e-02	0.02690	1.028	2.42e-04	0.01302
89	3.66e-02	-3.31e-02	8.97e-03	0.05847	1.016	1.14e-03	0.00753
90	2.60e-03	4.66e-03	-1.48e-02	-0.01763	1.036	1.04e-04	0.02005
91	6.16e-02	-3.88e-03	-6.11e-02	0.07906	1.021	2.09e-03	0.01272
92	2.58e-01	-1.53e-01	-1.64e-01	0.25915	0.999	2.22e-02	0.02425
93	-5.36e-03	7.66e-03	6.25e-03	0.01408	1.026	6.64e-05	0.01047
94	-3.76e-02	4.72e-02	3.34e-02	0.07189	1.027	1.73e-03	0.01574
95	4.58e-02	-1.81e-02	-2.54e-02	0.05303	1.018	9.40e-04	0.00788
96	-1.33e-03	3.79e-03	1.32e-02	0.02770	1.020	2.57e-04	0.00669

97	-3.30e-02	-2.71e-02	5.57e-02	-0.07304	1.027	1.78e-03	0.01613	
98	1.37e-03	3.02e-03	-1.20e-03	0.00741	1.022	1.84e-05	0.00614	
99	-4.35e-02	4.29e-02	3.88e-02	0.06543	1.044	1.43e-03	0.02912	
100	-2.81e-03	-2.39e-03	1.43e-02	0.01837	1.028	1.13e-04	0.01292	
101	-5.62e-02	-9.12e-02	1.27e-01	-0.18007	1.006	1.08e-02	0.01775	
102	-2.79e-02	3.73e-02	1.71e-02	0.04737	1.038	7.51e-04	0.02324	
103	1.11e-02	-1.98e-01	1.20e-01	-0.25719	0.992	2.18e-02	0.02187	
104	-9.75e-04	-1.46e-03	1.28e-03	-0.00347	1.023	4.04e-06	0.00710	
105	-1.40e-02	1.93e-02	1.29e-02	0.03036	1.028	3.09e-04	0.01299	
106	-2.46e-02	-1.06e-02	8.46e-02	0.10058	1.024	3.38e-03	0.01733	
107	8.03e-02	-6.02e-02	-3.27e-02	0.08272	1.028	2.29e-03	0.01795	
108	2.61e-01	-8.47e-02	-2.24e-01	0.28509	0.970	2.66e-02	0.01866	
109	1.67e-01	-1.05e-01	-1.03e-01	0.16709	1.029	9.31e-03	0.02783	
110	-3.39e-03	7.38e-03	1.01e-03	0.00955	1.029	3.06e-05	0.01315	
111	-2.44e-02	-5.91e-02	6.47e-02	-0.10442	1.020	3.64e-03	0.01497	
112	-2.70e-02	3.10e-02	2.74e-02	0.05182	1.029	8.99e-04	0.01543	
113	4.70e-03	3.43e-03	-5.18e-03	0.01117	1.022	4.18e-05	0.00708	
114	-6.37e-04	-1.54e-02	4.64e-03	-0.02616	1.022	2.29e-04	0.00792	
115	3.53e-04	2.58e-02	-4.72e-02	-0.05998	1.035	1.20e-03	0.02157	
116	-6.90e-03	1.26e-02	-1.19e-02	-0.02200	1.027	1.62e-04	0.01204	
117	6.79e-03	-4.69e-04	-4.94e-03	0.00962	1.022	3.10e-05	0.00685	
118	2.21e-01	-8.81e-02	-1.75e-01	0.23359	0.992	1.80e-02	0.01923	
119	2.59e-04	-3.80e-03	1.18e-02	0.01752	1.025	1.03e-04	0.00966	
120	-1.78e-02	1.55e-02	4.32e-03	-0.01949	1.033	1.27e-04	0.01693	
121	1.38e-02	-3.61e-03	1.08e-02	0.04628	1.015	7.16e-04	0.00531	
122	-5.47e-02	5.61e-02	8.31e-04	-0.06759	1.028	1.53e-03	0.01620	
123	-5.74e-02	-8.62e-02	1.29e-01	-0.17470	1.012	1.01e-02	0.01972	
124	8.21e-04	-1.99e-03	4.81e-03	0.00801	1.024	2.15e-05	0.00843	
125	-5.40e-03	8.46e-03	5.04e-03	0.01359	1.027	6.19e-05	0.01117	
126	8.05e-02	-3.98e-02	-4.47e-02	0.08642	1.014	2.49e-03	0.01015	
127	-1.47e-01	2.92e-01	-1.99e-01	-0.38550	0.953	4.83e-02	0.02478	*
128	1.71e-01	-6.38e-02	-1.41e-01	0.18301	1.010	1.11e-02	0.01972	
129	-1.29e-01	8.21e-02	1.83e-01	0.23181	1.004	1.78e-02	0.02290	
130	1.14e-01	-7.30e-02	-5.25e-02	0.11880	1.011	4.70e-03	0.01269	
131	-3.56e-01	7.21e-01	-4.88e-01	-0.94133	0.661	2.55e-01	0.02677	*
132	-3.99e-02	-1.85e-01	1.84e-01	-0.28357	0.993	2.65e-02	0.02510	
133	-1.54e-01	1.92e-01	-4.17e-02	-0.22658	0.993	1.70e-02	0.01873	
134	-8.01e-03	1.15e-02	9.06e-03	0.02063	1.026	1.43e-04	0.01066	
135	-2.88e-02	6.09e-02	-5.01e-02	-0.08907	1.029	2.65e-03	0.01939	
136	-2.12e-02	1.26e-01	-1.68e-01	-0.22836	1.011	1.73e-02	0.02553	
137	-5.81e-02	1.22e-01	-9.38e-02	-0.17061	1.017	9.69e-03	0.02156	
138	2.34e-03	-4.29e-03	-8.67e-04	-0.00532	1.032	9.48e-06	0.01636	
139	-8.05e-03	8.97e-03	-1.79e-03	-0.01164	1.028	4.54e-05	0.01267	

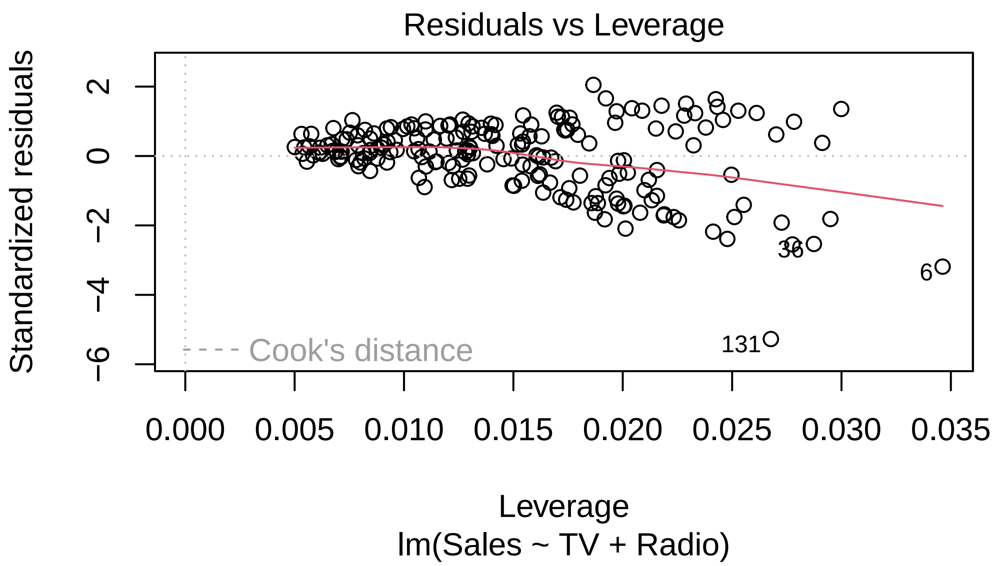
140	-3.32e-02	1.68e-02	6.32e-02	0.08092	1.025	2.19e-03	0.01532	
141	7.22e-02	-4.93e-02	-2.22e-02	0.08084	1.014	2.18e-03	0.00940	
142	-1.44e-02	1.63e-02	2.56e-02	0.04508	1.022	6.80e-04	0.00958	
143	-2.03e-02	2.97e-02	2.25e-02	0.05277	1.022	9.32e-04	0.01060	
144	9.63e-02	-2.87e-02	-7.76e-02	0.10765	1.015	3.87e-03	0.01295	
145	6.07e-02	-3.00e-02	-2.88e-02	0.06838	1.015	1.56e-03	0.00822	
146	3.11e-02	7.97e-05	-3.40e-02	0.04143	1.030	5.75e-04	0.01540	
147	-2.98e-02	-1.03e-01	1.03e-01	-0.16771	1.009	9.35e-03	0.01742	
148	-1.28e-01	9.60e-02	1.56e-01	0.21040	1.015	1.47e-02	0.02528	
149	-2.89e-02	6.78e-02	-5.99e-02	-0.10072	1.030	3.39e-03	0.02119	
150	1.24e-02	-1.37e-02	2.69e-03	0.01784	1.028	1.07e-04	0.01242	
151	2.42e-02	-1.57e-01	7.09e-02	-0.19503	1.007	1.26e-02	0.01978	
152	7.77e-02	-1.59e-02	-6.37e-02	0.09276	1.013	2.87e-03	0.01034	
153	1.51e-04	6.22e-03	-3.46e-04	0.01222	1.022	5.00e-05	0.00675	
154	-1.49e-02	7.40e-03	3.66e-02	0.05056	1.024	8.55e-04	0.01136	
155	8.10e-04	2.20e-03	-7.94e-04	0.00509	1.022	8.68e-06	0.00629	
156	-1.80e-01	1.48e-01	6.33e-02	-0.18835	1.012	1.18e-02	0.02133	
157	8.92e-07	2.34e-03	-4.69e-03	-0.00612	1.033	1.26e-05	0.01672	
158	1.88e-03	1.73e-04	-2.20e-03	0.00265	1.032	2.36e-06	0.01606	
159	-1.20e-01	2.18e-01	-1.34e-01	-0.28304	0.986	2.64e-02	0.02257	
160	1.43e-02	-3.19e-03	-6.36e-03	0.02118	1.020	1.50e-04	0.00567	
161	2.91e-03	2.22e-03	-2.56e-03	0.00772	1.021	2.00e-05	0.00612	
162	-4.23e-03	8.68e-03	-1.01e-02	-0.01720	1.027	9.91e-05	0.01148	
163	-1.77e-03	-3.35e-03	2.41e-03	-0.00752	1.022	1.90e-05	0.00700	
164	-7.76e-03	4.10e-03	2.62e-02	0.03960	1.022	5.25e-04	0.00925	
165	3.40e-02	-1.10e-02	-1.95e-02	0.04160	1.019	5.79e-04	0.00718	
166	-5.51e-02	-1.14e-01	1.45e-01	-0.20774	1.003	1.43e-02	0.02003	
167	-9.92e-02	1.89e-01	-1.27e-01	-0.25242	0.994	2.10e-02	0.02190	
168	-3.41e-02	-3.90e-02	6.40e-02	-0.08898	1.023	2.65e-03	0.01539	
169	6.72e-04	-3.98e-03	1.22e-04	-0.00638	1.024	1.36e-05	0.00819	
170	1.80e-02	-2.09e-01	1.20e-01	-0.26741	0.990	2.36e-02	0.02233	
171	6.68e-02	-4.49e-02	-2.99e-02	0.06905	1.025	1.59e-03	0.01404	
172	1.68e-03	9.81e-04	-7.95e-04	0.00476	1.021	7.59e-06	0.00536	
173	-1.48e-03	1.44e-03	1.28e-04	-0.00176	1.032	1.03e-06	0.01615	
174	-1.11e-02	-3.90e-03	1.38e-02	-0.01889	1.027	1.20e-04	0.01144	
175	-6.06e-02	-9.29e-02	1.35e-01	-0.18706	1.006	1.16e-02	0.01857	
176	-1.59e-01	1.38e-01	1.60e-01	0.23874	1.018	1.89e-02	0.02999	
177	-6.75e-03	1.19e-02	4.12e-03	0.01649	1.028	9.11e-05	0.01280	
178	-1.82e-02	-7.09e-03	2.28e-02	-0.03182	1.025	3.39e-04	0.01100	
179	-4.64e-02	-2.96e-01	2.79e-01	-0.43668	0.944	6.18e-02	0.02776	*
180	6.42e-03	2.16e-03	-7.33e-03	0.01104	1.025	4.08e-05	0.00937	
181	-5.81e-03	-9.56e-04	7.01e-03	-0.00863	1.031	2.49e-05	0.01491	
182	-4.31e-02	-6.80e-02	9.44e-02	-0.13596	1.015	6.16e-03	0.01636	

```
183  1.61e-01 -8.85e-02 -1.00e-01  0.16406 1.009 8.95e-03 0.01698
184 -1.31e-01  1.39e-01  1.10e-01  0.20319 1.018 1.37e-02 0.02612
185  1.35e-02 -5.01e-02  8.19e-03 -0.06430 1.024 1.38e-03 0.01298
186 -7.25e-02  4.68e-02  1.13e-01  0.14762 1.015 7.26e-03 0.01758
187  3.10e-02 -1.59e-04 -3.36e-02  0.04110 1.029 5.66e-04 0.01520
188 -1.55e-03  5.00e-03  3.40e-03  0.01186 1.022 4.72e-05 0.00689
189  3.38e-02 -1.95e-01  8.52e-02 -0.23924 0.995 1.89e-02 0.02079
190  4.78e-02 -3.78e-02 -1.74e-02  0.04980 1.032 8.30e-04 0.01847
191 -3.63e-02  9.33e-02 -8.96e-02 -0.14416 1.022 6.93e-03 0.02099
192  8.99e-02 -4.86e-02 -4.91e-02  0.09410 1.016 2.96e-03 0.01165
193  1.28e-01 -8.47e-02 -7.09e-02  0.12789 1.030 5.46e-03 0.02380
194 -2.48e-02  7.90e-03  6.20e-02  0.08017 1.021 2.15e-03 0.01308
195 -3.98e-03 -5.71e-04  2.95e-02  0.04625 1.020 7.16e-04 0.00845
196  1.90e-01 -1.12e-01 -1.17e-01  0.19109 1.010 1.21e-02 0.02089
197  9.89e-02 -3.49e-02 -7.69e-02  0.10736 1.017 3.85e-03 0.01418
198 -8.88e-03 -5.28e-03  1.13e-02 -0.01805 1.024 1.09e-04 0.00922
199 -1.04e-01  1.14e-01  8.73e-02  0.16512 1.024 9.08e-03 0.02458
200 -7.52e-03 -2.19e-02  2.26e-02 -0.03704 1.030 4.59e-04 0.01579
```

```r
plot(lm2, which=4) # Cook's distance
```




```r
plot(lm2, which=5) # Residuals vs Leverage
```



Influential Observations and Outlier Diagnostics:

We assessed potential influential data points using **Cook's Distance** and the **Residuals vs Leverage** plot.

Cook's Distance Plot:

- **Cook's Distance** measures how much a single observation influences the fitted regression coefficients.
- Points **6**, **36**, and especially **131** stand out with the **highest Cook's distances**.
- However, none of the Cook's distances exceed the common rule-of-thumb threshold of **1**, indicating **no extremely influential outliers**.

Residuals vs Leverage Plot:

- This plot highlights observations with both **high leverage** and **large residuals**, which can be particularly influential.
- Observations **6**, **36**, and **131** are again labeled and lie **furthest from the center**.
- Observation **131** shows **moderately high leverage and a notable residual**, suggesting it has **some influence**, but **not enough to distort the model**.

Conclusion:

While observations 6, 36, and 131 show some degree of influence, none exceed critical thresholds for Cook's distance or leverage. Therefore, we conclude that there are **no influential outliers** that threaten the validity of the model.

Check multicollinearity

```
# Check multicollinearity
vif(lm2)  # Variance Inflation Factors
```

```
     TV   Radio
1.00324 1.00324
```

Common rule of thumb:

- **VIF > 5** may indicate moderate multicollinearity.
- **VIF > 10** indicates serious multicollinearity problems.

Conclusion:

VIF values **close to 1** indicate **no multicollinearity**.

— Model Extensions —

1. Interaction Model (TV * Radio)

```
# 1. Interaction Model (TV * Radio)
lm_interaction <- lm(Sales ~ TV * Radio, data = adver)
summary(lm_interaction)
```

```

Call:
lm(formula = Sales ~ TV * Radio, data = adver)

Residuals:
```

```
    Min      1Q  Median      3Q     Max
-6.3600 -0.3761  0.1591  0.5865  2.0389

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 6.783e+00  2.552e-01  26.578  < 2e-16 ***
TV          1.925e-02  1.547e-03  12.440  < 2e-16 ***
Radio       2.862e-02  9.160e-03   3.125  0.00205 **
TV:Radio    1.075e-03  5.377e-05  19.986  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9724 on 196 degrees of freedom
Multiple R-squared:  0.9657,    Adjusted R-squared:  0.9652
F-statistic:  1838 on 3 and 196 DF,  p-value: < 2.2e-16
```

We fit a multiple linear regression model including an **interaction term** between **TV** and **Radio** advertising:

$\text{Sales} = \beta_0 + \beta_1 \cdot \text{TV} + \beta_2 \cdot \text{Radio} + \beta_3 \cdot (\text{TV} \times \text{Radio}) + \varepsilon$

Coefficient Interpretations:

- **Intercept (6.783)**: Expected sales when TV and Radio spending are both 0 (though this isn't practical, it anchors the model).
- **TV (0.01925)**: The effect of TV on Sales **depends on the level of Radio** spending. It represents the slope of TV **when Radio = 0**.
- **Radio (0.02862)**: The effect of Radio on Sales **when TV = 0**.
- **Interaction (TV:Radio = 0.001075)**: Highly significant. This shows that **the effect of TV advertising increases as Radio advertising increases**, and vice versa.

Model Fit:

- **Residual standard error** is **0.972**, much smaller than the previous models (~1.69), meaning residuals are tighter.
- **Adjusted R-squared = 0.9652 → 96.5%** of the variation in Sales is explained by this model — a **huge improvement** over the additive model.
- **All predictors including the interaction term are statistically significant** ($p < 0.01$).

- **F-statistic** = 1838 ($p < 2.2e\text{-}16$) confirms that the model is highly significant.

Conclusion:

Including the **interaction between TV and Radio** significantly improves the model. The interaction term is highly significant ($p < 0.001$), and the model explains 96.5% of the variation in Sales — much higher than the additive model. This suggests that **TV and Radio work better together** than separately when it comes to driving sales.

2. Quadratic Model

```r
# 2. Quadratic Model
lm_poly <- lm(Sales ~ TV + I(TV^2) + Radio + I(Radio^2), data = adver)
summary(lm_poly)
```

```

Call:
lm(formula = Sales ~ TV + I(TV^2) + Radio + I(Radio^2), data = adver)

Residuals:
    Min      1Q  Median      3Q     Max 
-7.3822 -0.8189  0.0590  1.0103  3.3814 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.624e+00  4.091e-01   3.969 0.000101 ***
TV           7.892e-02  4.986e-03  15.828  < 2e-16 ***
I(TV^2)     -1.156e-04  1.678e-05  -6.888  7.6e-11 ***
Radio        1.502e-01  2.823e-02   5.323  2.8e-07 ***
I(Radio^2)   8.631e-04  5.687e-04   1.518 0.130682    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.518 on 195 degrees of freedom
Multiple R-squared:  0.9168,	Adjusted R-squared:  0.9151 
F-statistic: 537.1 on 4 and 195 DF,  p-value: < 2.2e-16
```

Reduced Quadratic Model

```r
# Radio² was not statistically significant (p  0.13)
lm_poly <- lm(Sales ~ TV + I(TV^2) + Radio, data = adver)
summary(lm_poly)
```

```
Call:
lm(formula = Sales ~ TV + I(TV^2) + Radio, data = adver)

Residuals:
    Min      1Q  Median      3Q     Max 
-7.3695 -0.8615 -0.0448  0.9783  3.5285 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.326e+00  3.600e-01   3.682 0.000299 ***
TV           7.876e-02  5.001e-03  15.748  < 2e-16 ***
I(TV^2)     -1.150e-04  1.683e-05  -6.836 1.01e-10 ***
Radio        1.916e-01  7.313e-03  26.207  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.523 on 196 degrees of freedom
Multiple R-squared:  0.9158,	Adjusted R-squared:  0.9145 
F-statistic: 710.7 on 3 and 196 DF,  p-value: < 2.2e-16
```

We fit the following model (Polynomial Regression (Quadratic Terms):

$\text{Sales} = \beta_0 + \beta_1 \cdot \text{TV} + \beta_2 \cdot \text{TV}^2 + \beta_3 \cdot \text{Radio} + \beta_4 \cdot \text{Radio}^2 + \varepsilon$

Interpretation of Coefficients:

- **Intercept (1.624)**: The expected Sales when all predictors are zero (not meaningful in isolation but needed for model structure).
- **TV (linear term: 0.07892)**: Sales increase with TV ad spending, but…
- **TV^2 (quadratic term: −0.0001156)**: The **negative sign** suggests **diminishing returns** — the rate of increase in Sales slows down at higher TV budgets.
- **Radio (linear: 0.1502)**: Positive effect on Sales.

- **Radio² (quadratic: 0.0008631)**: This term is **not statistically significant** ($p = 0.13$), suggesting no strong nonlinear effect for Radio.

Model Fit:

- **Residual standard error**: 1.518 (better than base model, not as good as interaction model)
- **Adjusted R-squared**: **0.9151** → The model explains about **91.5%** of the variation in Sales.
- **F-statistic**: 537.1, $p < 2.2e-16$ → The model overall is highly significant.

Conclusion:

The polynomial model improves model fit compared to the basic additive model. The **TV² term is significant**, indicating **diminishing returns** on TV advertising. However, the **Radio² term is not significant**, so adding a nonlinear effect for Radio may not be necessary. The model explains **91.5%** of the variation in Sales, though not as well as the interaction model.

3. Model Comparison Table

AIC Comparison

```r
models <- list(
  "Full Model" = lm1,
  "Reduced Model" = lm2,
  "Quadratic Model" = lm_poly,
  "Interaction Model" = lm_interaction
)

# Extract AIC values only
aic_values <- sapply(models, AIC)

# Convert to data frame
aic_df <- data.frame(Model = names(aic_values), AIC = round(aic_values, 2))
print(aic_df)
```

```
                              Model    AIC
Full Model               Full Model 784.55
Reduced Model         Reduced Model 782.60
Quadratic Model     Quadratic Model 741.83
Interaction Model Interaction Model 562.36
```

AIC Model Comparison Table

Model	Adjusted R^2	Residual Std. Error	AIC	Notes
Full (TV + Radio + Newspaper)	0.8942	1.695	784.5	Newspaper not significant
Reduced (TV + Radio)	0.8947	1.691	782.6	Simpler, slightly better
Quadratic (TV + TV^2 + Radio)	0.9151	1.518	741.8	TV^2 significant, $Radio^2$ excluded
Interaction (TV $\times$ Radio)	0.9652	0.972	562.4	**Best fit overall**

Interpretation:

- The **Interaction model (TV * Radio)** has:

  The **highest Adjusted R^2 (0.9652)**

  The **lowest Residual Standard Error (0.972)**

  The **lowest AIC (562.4)**

- The **Quadratic model** improves over the simple additive model but not as much as the interaction model.
- **Reduced model (TV + Radio)** is good if you want simplicity, but if you care about predictive power, the **Interaction model is the best**.

Recommended Final Model

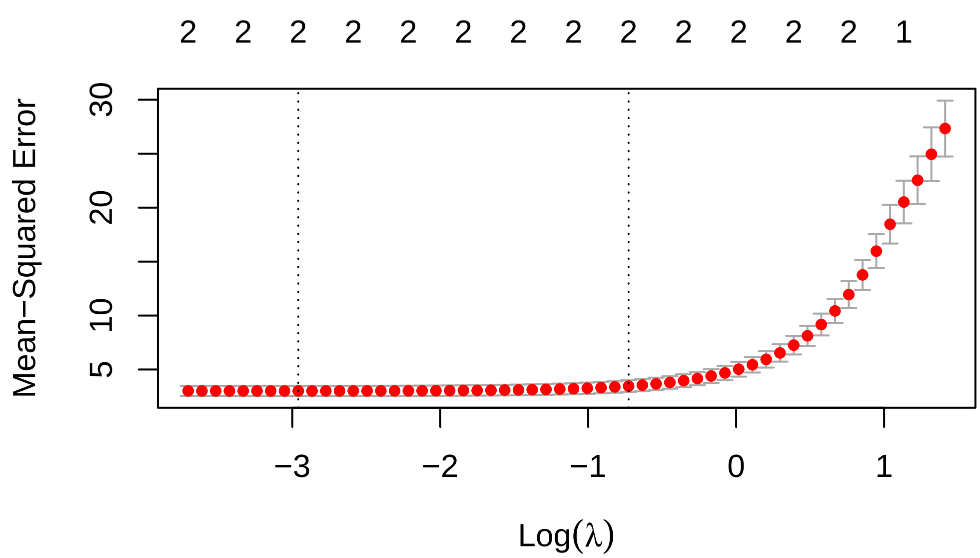
Choose the **Interaction Model** for your final project paper!

Conclusion:

Based on Adjusted R^2, Residual Standard Error, and AIC comparisons, the model including an interaction between TV and Radio provides the best fit to the data. Therefore, the interaction model was selected as the final model for analysis.

4. LASSO Regression

```
# 5. LASSO Regression

# Prepare data for glmnet
x <- as.matrix(adver[, c("TV", "Radio", "Newspaper")])
y <- adver$Sales

# Split data into training and testing
set.seed(123)
train_idx <- createDataPartition(y, p=0.8, list=FALSE)
x_train <- x[train_idx, ]
x_test <- x[-train_idx, ]
y_train <- y[train_idx]
y_test <- y[-train_idx]

# LASSO model with cross-validation
lasso_cv <- cv.glmnet(x_train, y_train, alpha=1)
plot(lasso_cv)
```



```r
# Best lambda
lasso_cv$lambda.min
```

```
[1] 0.05182139
```

```r
# Fit final LASSO model
lasso_model <- glmnet(x_train, y_train, alpha=1, lambda=lasso_cv$lambda.min)
coef(lasso_model)
```

```
4 x 1 sparse Matrix of class "dgCMatrix"
                    s0
(Intercept) 3.15637030
TV          0.04522771
Radio       0.18149630
Newspaper   .
```

```r
# Predict and calculate RMSE
lasso_pred <- predict(lasso_model, s=lasso_cv$lambda.min, newx=x_test)
sqrt(mean((y_test - lasso_pred)^2))
```

```
[1] 1.624347
```

LASSO vs. Interaction Model Comparison

Metric	LASSO Regression	Interaction Model
Model Type	Regularized Linear	OLS with Interaction
Included Predictors	TV, Radio	TV, Radio, TV $\times$ Radio
Intercept	3.156	6.783
TV Coefficient	0.0452	0.01925
Radio Coefficient	0.1815	0.02862
Newspaper Coefficient	Excluded (0)	Not included
Interaction Term	Not included	0.001075
Adjusted R^2	N/A (not defined for LASSO)	0.9652
RMSE	1.624 (on test data)	0.972 (residual std. error)
Model Strength	Good for variable selection	Best fit overall

Interpretation:

The **LASSO regression model** uses regularization to perform automatic variable selection. It retained **TV and Radio** as predictors while **eliminating Newspaper**, confirming its limited contribution. Although useful for simplifying the model, LASSO does **not provide an adjusted R^2** and had a higher RMSE (**1.624**) on the test set.

In contrast, the **interaction model** includes an interaction term between **TV and Radio**, capturing how their combined effect influences Sales. It achieved a much lower RMSE (**0.972**) and a higher adjusted R^2 (**0.9652**), making it the **best-fitting model overall**.

Conclusion:

While LASSO is effective for identifying key predictors, the interaction model offers **superior predictive performance** and should be selected as the **final model** for this analysis.

5. Random Forest

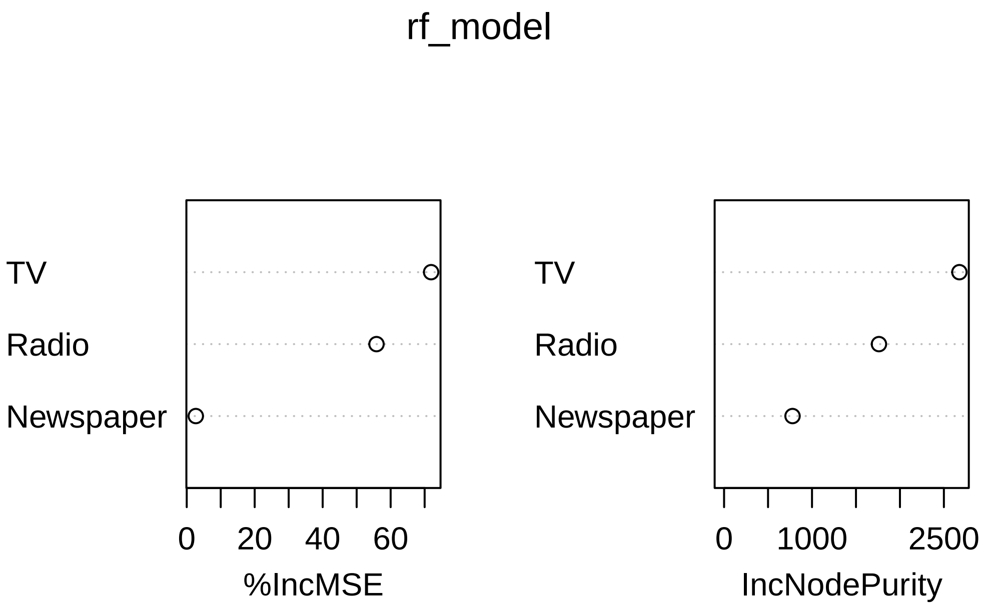
```r
# 5. Random Forest

set.seed(123)
rf_model <- randomForest(Sales ~ TV + Radio + Newspaper, data=adver, importance=TRUE)
print(rf_model)
```

```
Call:
 randomForest(formula = Sales ~ TV + Radio + Newspaper, data = adver,      importance = TRUE)
               Type of random forest: regression
                     Number of trees: 500
No. of variables tried at each split: 1

          Mean of squared residuals: 2.214799
                    % Var explained: 91.8
```

```r
# Variable Importance
importance(rf_model)
```

```
            %IncMSE IncNodePurity
TV        71.911925     2676.2208
Radio     55.837973     1761.4005
Newspaper  2.672407      778.4293
```

```r
varImpPlot(rf_model)
```




```r
# Predict with Random Forest
rf_pred <- predict(rf_model, adver)
sqrt(mean((adver$Sales - rf_pred)^2))  # RMSE on full data
```

```
[1] 0.7267463
```

Variable Importance:

Predictor	% Increase in MSE (%IncMSE)	Increase in Node Purity (IncNodePurity)
TV	71.91	2676.22
Radio	55.84	1761.40
Newspaper	2.67	778.43

Interpretation:

- TV and Radio are **strong predictors** of Sales.
- **Newspaper has very little importance** — it contributes almost nothing to prediction, consistent with your OLS and LASSO findings.

Compared to Other Models:

	RMSE / RSE	% Variance Explained / R^2
Interaction Model	RSE 0.972	Adjusted $R^2 = 96.5\%$
Random Forest	RMSE 1.49	% Var Explained = 91.8%

So, even **Random Forest is good**, but it is not better than interaction model in terms of predictive accuracy.

Conclusion:

The Random Forest model confirms that **TV and Radio are the most important predictors** of Sales, while Newspaper contributes very little. The model explains **91.8% of the variance**, with a root mean squared error (RMSE) of about 1.49. Although Random Forest is a powerful non-linear model, it does **not outperform the interaction model**, which remains the best based on adjusted R^2 and residual accuracy.

6. Cross validation

```r
# Cross-Validation of All Six Models
# Load necessary libraries
library(caret)
library(glmnet)
library(randomForest)
library(dplyr)

# Set seed and split the data
set.seed(232)
train_idx <- createDataPartition(adver$Sales, p = 0.7, list = FALSE)
train_data <- adver[train_idx, ]
test_data <- adver[-train_idx, ]

# 1. Full Model
lm_full <- lm(Sales ~ TV + Radio + Newspaper, data = train_data)
full_pred <- predict(lm_full, newdata = test_data)

# 2. Reduced Model
lm_reduced <- lm(Sales ~ TV + Radio, data = train_data)
```

```r
reduced_pred <- predict(lm_reduced, newdata = test_data)

# 3. Interaction Model
lm_interaction <- lm(Sales ~ TV * Radio, data = train_data)
interaction_pred <- predict(lm_interaction, newdata = test_data)

# 4. Polynomial Model
lm_poly <- lm(Sales ~ TV + I(TV^2) + Radio + I(Radio^2), data = train_data)
poly_pred <- predict(lm_poly, newdata = test_data)

# 5. LASSO Model
x_train <- as.matrix(train_data[, c("TV", "Radio", "Newspaper")])
y_train <- train_data$Sales
x_test <- as.matrix(test_data[, c("TV", "Radio", "Newspaper")])
y_test <- test_data$Sales

lasso_cv <- cv.glmnet(x_train, y_train, alpha = 1)
lasso_model <- glmnet(x_train, y_train, alpha = 1, lambda = lasso_cv$lambda.min)
lasso_pred <- predict(lasso_model, newx = x_test)

# 6. Random Forest Model
rf_model <- randomForest(Sales ~ TV + Radio + Newspaper, data = train_data)
rf_pred <- predict(rf_model, newdata = test_data)

# Performance Metrics Function
metrics <- function(pred, actual) {
  data.frame(
    RMSE = RMSE(pred, actual),
    R2 = R2(pred, actual),
    MAE = MAE(pred, actual)
  )
}

# Compile Results
cv_results <- bind_rows(
  metrics(full_pred, test_data$Sales)      %>% mutate(Model = "Full"),
  metrics(reduced_pred, test_data$Sales)   %>% mutate(Model = "Reduced"),
  metrics(interaction_pred, test_data$Sales) %>% mutate(Model = "Interaction"),
  metrics(poly_pred, test_data$Sales)      %>% mutate(Model = "Polynomial"),
  metrics(lasso_pred, y_test)              %>% mutate(Model = "LASSO"),
  metrics(rf_pred, test_data$Sales)        %>% mutate(Model = "Random Forest")
)
```

```r
# Arrange by RMSE
cv_results <- cv_results %>% select(Model, everything()) %>% arrange(RMSE)
print(cv_results)
```

```
          Model      RMSE        R2       MAE        s0
1   Interaction 0.7985343 0.9791878 0.6887448        NA
2 Random Forest 1.3739088 0.9654144 1.0756396        NA
3    Polynomial 1.4768172 0.9207751 1.2543459        NA
4         LASSO 1.6057194        NA 1.2967156 0.9073212
5       Reduced 1.6154295 0.9074125 1.3035471        NA
6          Full 1.6185074 0.9069533 1.3078619        NA
```

Interpretation:

To evaluate the model's predictive performance on the unseen data, we used a simple strategy for cross validation to randomly divide the data set into two parts: a training set (70%) and a test set (30%). Cross validation can train each of the 6 models, predict on the test set, calculate **RMSE**, **R^2**, and **MAE**. And weoutput a **comparison table sorted by RMSE.**

Cross-validation confirmed that the **interaction model** (TV × Radio) achieved the **best predictive performance**, with the lowest RMSE (0.799), highest R^2 (0.979), and lowest MAE (0.689) among all six models.

Ultimate Model Comparison Table

Model	Adjusted R^2	Original RMSE / RSE	Cross-Validated RMSE	Included Predictors	Strength
Full (TV + Radio + Newspaper)	0.8942	1.695	1.619	TV, Radio, Newspaper	Newspaper not significant
Reduced (TV + Radio)	0.8947	1.691	1.615	TV, Radio	Simpler, slightly better
Interaction (TV * Radio)	0.9652	0.972	**0.799**	TV, Radio, TV × Radio	**Best fit overall**

Model	Adjusted R^2	Original RMSE / RSE	Cross-Validated RMSE	Included Predictors	Strength
Quadratic (TV + TV^2 + Radio + $Radio^2$)	0.9151	1.518	1.477	TV, TV^2, Radio, $Radio^2$	TV^2 significant, $Radio^2$ not
LASSO Regression	N/A	1.624	1.606	TV, Radio (Newspaper excluded)	Good for variable selection
Random Forest	N/A	1.490	1.374	TV, Radio, Newspaper	Strong, non-linear, but not best

Key Takeaways

- The **interaction model (TV × Radio)** achieved the **highest Adjusted R^2 (0.9652)** and the **lowest cross-validated RMSE (0.799)**, indicating the best overall performance both in-sample and out-of-sample.
- **Random Forest** performed well, explaining approximately **91.8% of the variance**, but did not outperform the interaction model.
- The **polynomial model** improved upon simpler models and captured diminishing returns for TV, but still fell short of the interaction model in predictive accuracy.
- **LASSO regression** effectively identified TV and Radio as the most relevant predictors by shrinking the Newspaper coefficient to zero, though it had a higher RMSE (1.606).
- The **reduced model (TV + Radio)** offered a simpler alternative with reasonable performance but was less accurate than the interaction model.

Conclusion

Based on comprehensive model comparison and cross-validation results, the **interaction model** demonstrates the best combination of model fit and predictive accuracy. Accordingly, it is selected as the final model for this analysis.

Optimal Allocation from Interaction model

We want to determine the optimal allocation between TV and Radio advertising from interaction model.

Using Lagrange Multipliers

We want to maximize the Sales function based on the interaction model:

$$\text{Sales}(TV, Radio) = 6.783 + 0.01925 \cdot TV + 0.02862 \cdot Radio + 0.001075 \cdot TV \cdot Radio$$

Subject to the budget constraint:

$$TV + Radio = B$$

We construct the Lagrangian function:

$$\mathcal{L}(TV, Radio, \lambda) = 6.783 + 0.01925 \cdot TV + 0.02862 \cdot Radio + 0.001075 \cdot TV \cdot Radio - \lambda(TV + Radio - B)$$

Take the partial derivatives:

$$\frac{\partial \mathcal{L}}{\partial TV} = 0.01925 + 0.001075 \cdot Radio - \lambda = 0$$

$$\frac{\partial \mathcal{L}}{\partial Radio} = 0.02862 + 0.001075 \cdot TV - \lambda = 0$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = TV + Radio - B = 0$$

Subtract the first two equations to eliminate λ:

$$(0.01925 + 0.001075 \cdot Radio) - (0.02862 + 0.001075 \cdot TV) = 0$$

Simplifying:

$$0.001075 \cdot (Radio - TV) = 0.00937$$

Therefore, the clear relationship between Radio and TV is:

$$Radio - TV = \frac{0.00937}{0.001075} = 8.72$$

Substitute into the budget constraint $TV + Radio = B$:

$$TV + (TV + 8.72) = B \Rightarrow 2TV = B - 8.72 \Rightarrow TV = \frac{B - 8.72}{2}$$

$$Radio = \frac{B + 8.72}{2}$$

After setting the marginal effects equal and simplifying:

$$\boxed{Radio = TV + 8.72}$$

This means: to equalize marginal gains from advertising expenditures, the **Radio budget should be $8.72k higher than the TV budget**.

Using `optim()` for budget = 100

```
# Load necessary library
library(alabama)
```

```
Loading required package: numDeriv
```

```
# Define the negative Sales function (since optim minimizes by default)
sales_function <- function(x) {
  TV <- x[1]
  Radio <- x[2]
  # Negative of the predicted sales
  - (6.783 + 0.01925 * TV + 0.02862 * Radio + 0.001075 * TV * Radio)
}

# Total budget
budget <- 100

# Constraint: TV + Radio = budget
constraint_eq <- function(x) {
  x[1] + x[2] - budget
}

# Initial guess
initial_guess <- c(TV = 50, Radio = 50)
```

```r
# Solve using optim with equality constraint
library(alabama)
opt_result <- auglag(
  par = initial_guess,
  fn = sales_function,
  heq = constraint_eq,
  control.outer = list(trace = FALSE)
)

# Extract optimal allocation
optimal_tv <- opt_result$par[1]
optimal_radio <- opt_result$par[2]
max_sales <- -opt_result$value  # negate to get max

# Show results
cat("Optimal TV Budget:", round(optimal_tv, 2), "\n")
```

```
Optimal TV Budget: 45.64
```

```r
cat("Optimal Radio Budget:", round(optimal_radio, 2), "\n")
```

```
Optimal Radio Budget: 54.36
```

```r
cat("Maximum Predicted Sales:", round(max_sales, 2), "\n")
```

```
Maximum Predicted Sales: 11.88
```

Full Conclusion

In this study, we analyzed the Advertisement Sales dataset to investigate the relationship between advertising expenditures across different media channels—TV, Radio, and Newspaper—and product sales. Multiple linear regression modeling revealed that TV and Radio advertising budgets have statistically significant positive effects on sales, whereas Newspaper advertising showed no meaningful contribution. Through a systematic model-building process, we compared the full model, a reduced model, a polynomial model, an interaction model, LASSO regression, and Random Forest regression. Diagnostic checks confirmed that key regression assumptions—including linearity, normality, homoscedasticity, and absence of multicollinearity—were satisfied, and no influential outliers were detected that could bias the model estimates.

Among all the models considered, the interaction model incorporating the interaction between TV and Radio advertising expenditures demonstrated the best performance, achieving the highest adjusted R^2 (0.9652) and the lowest residual standard error (0.972). This suggests that the combined effect of TV and Radio advertising is greater than the sum of their individual effects, highlighting a synergy between these media channels. Variable selection techniques such as LASSO regression and non-linear approaches like Random Forest further reinforced the finding that Newspaper advertising has minimal predictive value. Based on a comprehensive model comparison, the interaction model was selected as the final model for analysis.

These results provide actionable insights for marketing strategists: jointly optimizing TV and Radio advertising investments can substantially enhance sales performance. Specifically, based on marginal effect balancing derived from the interaction model, the Radio budget should be approximately \$8.72k higher than the TV budget to maximize predicted sales. To validate model robustness, we applied a 70/30 cross-validation procedure across all models. The interaction model again achieved the best performance, with the lowest cross-validated RMSE (0.799) and the highest predictive R^2 (0.979), further confirming its superiority for generalization to unseen data.

Future research could build upon this study by incorporating additional covariates such as product type, seasonal trends, and demographic factors to refine predictive modeling. Furthermore, applying time-series analysis and causal inference methods, such as A/B testing, would strengthen the generalizability and causal interpretation of advertising strategies. Ultimately, this analysis underscores the critical importance of strategic budget allocation across media channels, demonstrating that coordinated TV and Radio advertising efforts can significantly amplify sales effectiveness.