STAT 632 Project

Impact of Media Advertising on Sales Performance
A Comprehensive Analysis of TV, Radio, and Newspaper

Authors: David Teng , Hao Shen

Date: 05/01/2025

Introduction

Goals:

Analyze the impact of advertising budgets (TV, Radio, and Newspapers) on sales. Additionally, find an effective strategy for maximizing profit.

Statistical Method:

Multiple Linear Regression Model,
Reduce Model,
Quadratic Model,
Interaction Model (TV × Radio)

Data Description

The Advertisement Sales dataset is collected from the Kaggle website.

This dataset consists of **200** observations, and **5** variables.

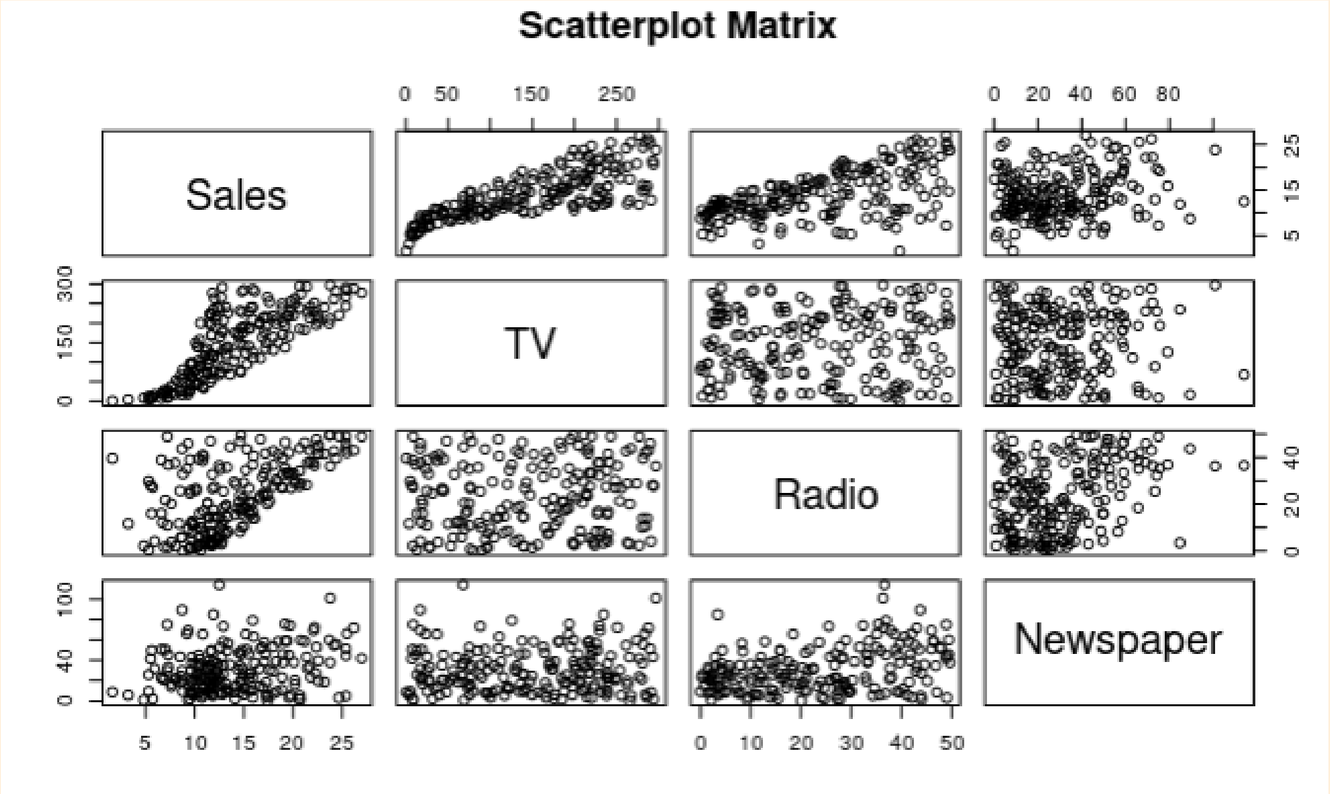
Predictor variables :

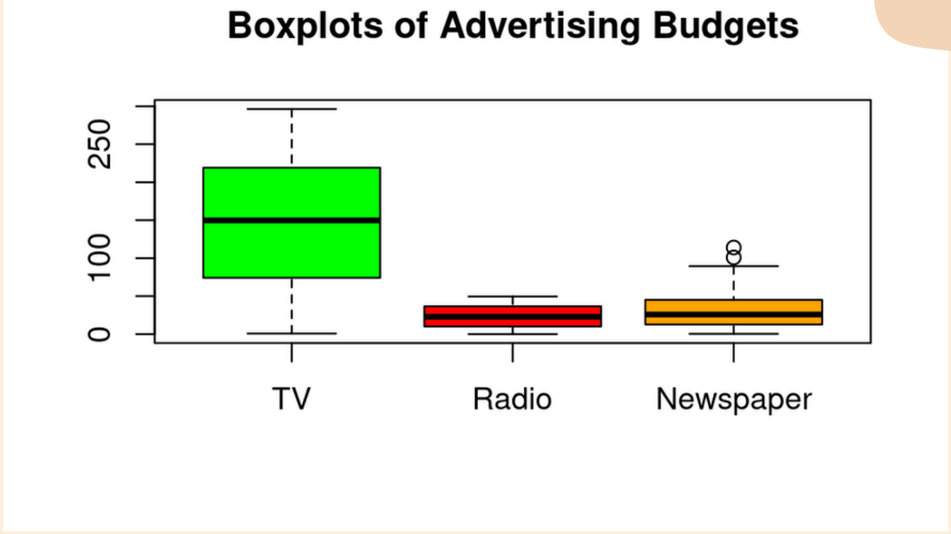
TV: Investment in TV advertising campaigns.
Radio: Investment in radio advertising campaigns.
Newspaper: Investment in newspaper advertising campaigns

Response variables:

Sales: The sales figures for the product (in thousands of **units**)

Data Description






```r
colSums(is.na(adver))
```

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

There are no missing values in our dataset.

Multiple linear regression

```
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
```

Multiple Linear Regression Results

- Goal: Assess the individual impact of TV, Radio, and Newspaper advertising on Sales.
- Significant Predictors:
  TV ($p < 0.05$)
  Radio ($p < 0.05$)
- Not Significant:
  Newspaper ($p = 0.822$)
- Adjusted $R^2 = 0.8958$ → a strong fit.

Reduced model

```
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
```

Reduced Model → Excluding **Newspaper**

- Reason: Newspaper was not statistically significant (p = 0.822)

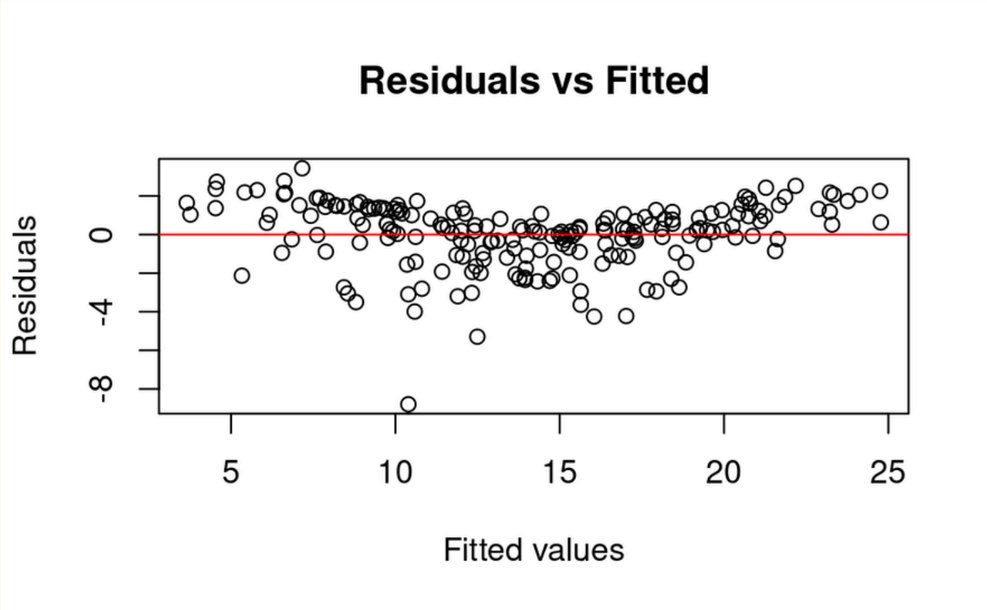
- Adjusted R^2 (Reduced Model): **0.8947**
- Adjusted R^2 (Full model): 0.8942
  → Reduced Model is a **better fit**

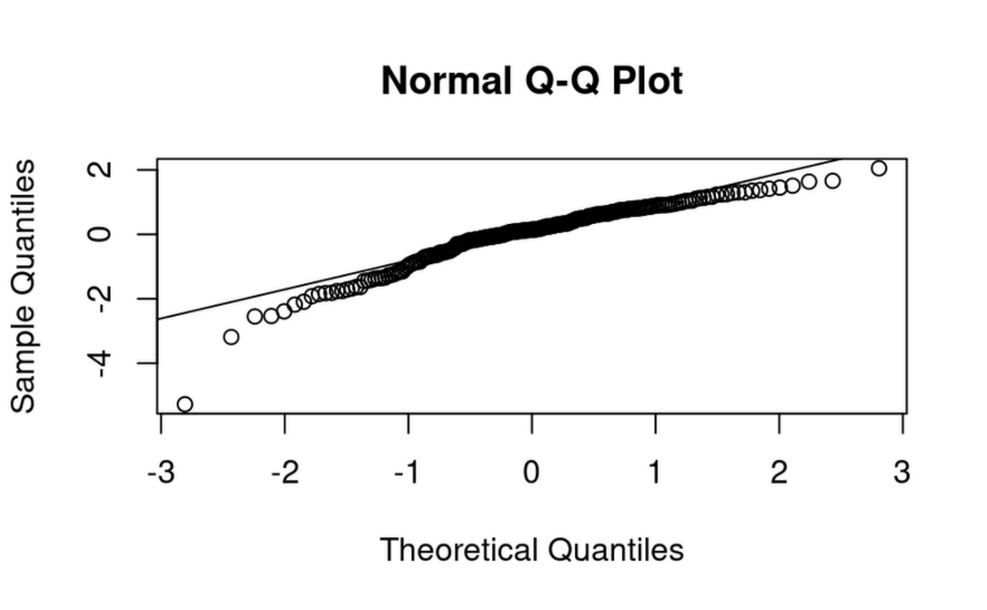
```
Analysis of Variance Table

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
```

- P-value > 0.05 → Fail to reject the null hypothesis
- Newspaper does not significantly improve the model
- Proceed with reduced model (without Newspaper)

Model Assumptions Verification

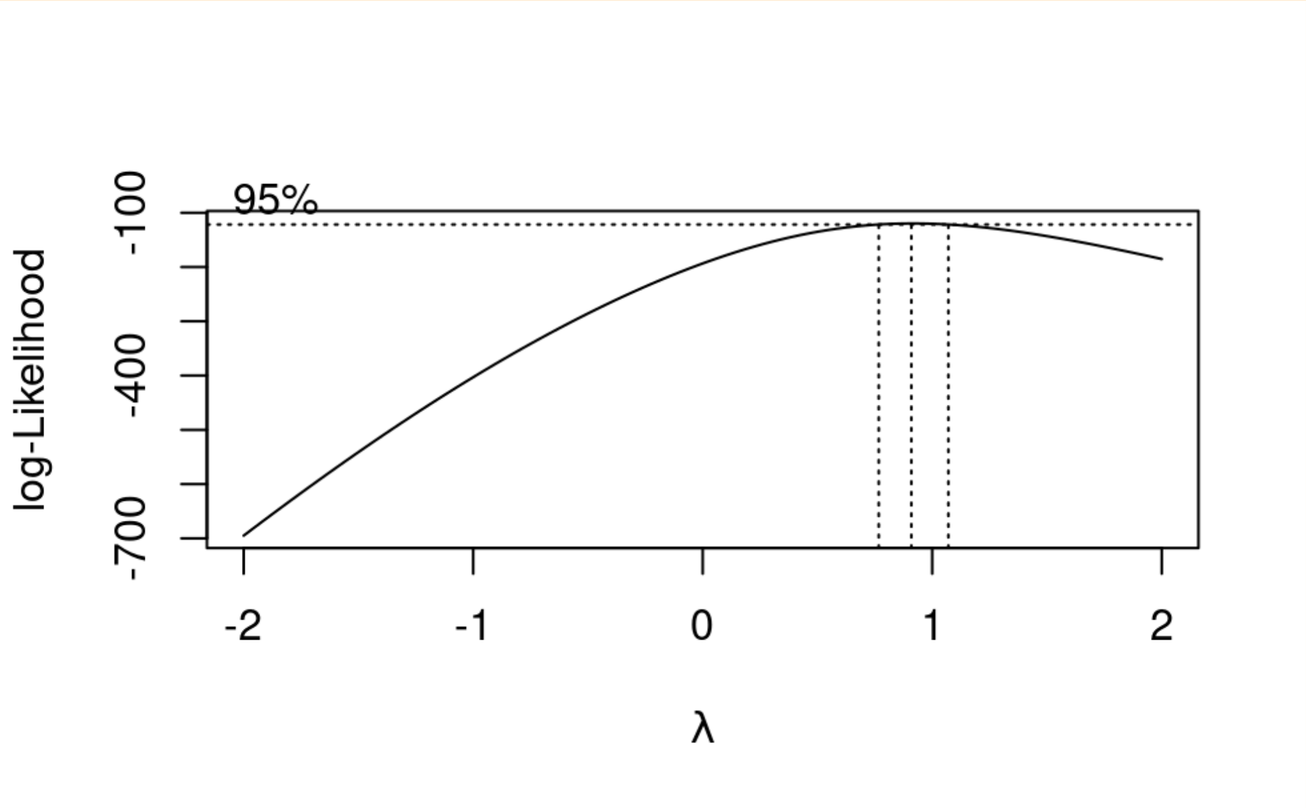




Linearity and **constant variance** are satisfied

Normality is mildly satisfied, with minor deviations at the tails.

Model Assumptions Verification



Since Q-Q plot shows slight non-linearity and mild skewness, we want to check if Box-Cox transformation is needed.

The plot shows the optimal λ is close to **1**

Therefore, **no transformation of the response variable is needed**

Interaction model(TV × Radio)

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

Residuals:
    Min      1Q  Median      3Q     Max
-5.9139 -0.3535  0.1713  0.5706  1.9917

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         6.586e+00  4.798e-01  13.726  < 2e-16 ***
TV                  1.997e-02  2.801e-03   7.128 2.01e-11 ***
Radio               1.928e-02  1.688e-02   1.142    0.255
Newspaper           1.322e-02  1.774e-02   0.745    0.457
TV:Radio            1.150e-03  1.004e-04  11.447  < 2e-16 ***
TV:Newspaper       -6.036e-05  9.601e-05  -0.629    0.530
Radio:Newspaper     1.013e-05  4.977e-04   0.020    0.984
TV:Radio:Newspaper -7.067e-07  2.778e-06  -0.254    0.799
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9695 on 192 degrees of freedom
Multiple R-squared:  0.9666,    Adjusted R-squared:  0.9654
F-statistic: 793.4 on 7 and 192 DF,  p-value: < 2.2e-16
```

	Model <chr>	AIC <dbl>
Full Model	Full Model	784.55
Reduced Model	Reduced Model	782.60
Quadratic Model	Quadratic Model	741.83
Interaction Model	Interaction Model	562.36

IModel Overview:

We created a multiple linear regression model to analyz impact of TV, Radio, and Newspaper advertising, includin interactions, on sales.

Key Findings:

1. TV Advertising: TV has a significant positive effect on sales ($p < 2e-16$), indicating a strong contribution.
2. Radio & Newspaper Advertising: Neither Radio ($p = 0.255$) nor Newspaper ($p = 0.457$) showed significant effects on sales.
3. Interactions: The TV × Radio interaction is significant ($p < 2e-16$), showing a synergistic effect. Other interactions, including TV × Newspaper and Radio × Newspaper, were not significant.

Model Performance:

- Adjusted R^2 = 96.54% and a high F-statistic (793.4) indicate a strong model fit.

Conclusion:

TV advertising significantly boosts sales, especially when combined with Radio. However, Radio and Newspaper have limited impact on sales, and their interactions do not enhance the model.

Quadratic model

```
Call:
lm(formula = Sales ~ TV + I(TV^2) + Radio + I(Radio^2) + New
    I(Newspaper^2) + I(TV^3) + I(TV^4) + I(TV^5), data = adv

Residuals:
    Min      1Q  Median      3Q     Max
-4.6928 -0.6546  0.1337  0.7461  3.8465

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    -1.460e+00  6.749e-01  -2.164  0.03173 *
TV              2.669e-01  4.326e-02   6.169 4.06e-09 ***
I(TV^2)        -3.254e-03  9.064e-04  -3.590  0.00042 ***
Radio           1.579e-01  2.609e-02   6.054 7.41e-09 ***
I(Radio^2)      7.851e-04  5.260e-04   1.493  0.13717
Newspaper       1.081e-02  1.352e-02   0.799  0.42500
I(Newspaper^2) -1.604e-04  1.588e-04  -1.010  0.31379
I(TV^3)         2.187e-05  7.731e-06   2.829  0.00517 **
I(TV^4)        -6.954e-08  2.871e-08  -2.422  0.01636 *
I(TV^5)         8.376e-11  3.853e-11   2.174  0.03096 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’

Residual standard error: 1.396 on 190 degrees of freedom
Multiple R-squared:  0.9314,    Adjusted R-squared:  0.9282
```

We applied a polynomial regression model to assess the effects of TV, Radio, and Newspaper advertising on sales, including their higher-order terms.

Key Findings:

1. TV Advertising: TV has a significant positive impact on sales, with higher-order terms (TV^2, TV^3, TV^4, TV^5) showing diminishing returns as the investment increases.
2. Radio Advertising: Radio also has a significant positive effect on sales, but its quadratic term is not significant.
3. Newspaper Advertising: Newspaper and its quadratic term were not significant, indicating a minimal impact on sales.

Model Performance:

- Adjusted R^2 = 92.82% and a high F-statistic (286.7) demonstrate a good fit of the model.

Conclusion:

TV and Radio advertising significantly boost sales, with diminishing returns for TV as investment increases. Newspaper advertising has limited impact and can be excluded to simplify the model.

Final model

```
Call:
lm(formula = Sales ~ TV + I(TV^2) + Radio + TV * Radio + I(
    I(TV^4) + I(TV^5), data = adver)

Residuals:
     Min       1Q   Median       3Q      Max
-2.98313 -0.20339  0.00319  0.20721  1.38233

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.713e+00  2.282e-01  11.888  < 2e-16 ***
TV           1.996e-01  1.402e-02  14.240  < 2e-16 ***
I(TV^2)     -2.528e-03  2.920e-04  -8.658 1.94e-15 ***
Radio        4.300e-02  4.297e-03  10.008  < 2e-16 ***
I(TV^3)      1.633e-05  2.489e-06   6.560 4.87e-10 ***
I(TV^4)     -4.973e-08  9.236e-09  -5.385 2.10e-07 ***
I(TV^5)      5.702e-11  1.239e-11   4.603 7.56e-06 ***
TV:Radio     1.025e-03  2.510e-05  40.834  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘

Residual standard error: 0.4503 on 192 degrees of freedom
Multiple R-squared:  0.9928,    Adjusted R-squared:  0.9925
F-statistic:  3778 on 7 and 192 DF,  p-value: < 2.2e-16
```

Final Model Summary:

- Adjusted R^2: 0.9925 (Excellent fit, explains 99.25% of sales variance)
- F-statistic: 3778 ($p < 2.2e-16$, highly significant model)
- Significant Variables:
  - TV has a strong positive effect on sales.
  - Radio positively impacts sales, with a significant TV:Radio interaction showing synergy.
  - Higher-order TV terms (TV^2, TV^3, TV^4, TV^5) indicate diminishing returns with increasing TV spending.
- Residual Standard Error: 0.4503 (Low, indicating a good fit)

Conclusion: The model fits well, with all terms significant and a very high explanatory power. The interaction between TV and Radio, along with nonlinear effects of TV, makes this model highly predictive for sales.

Cross validation

Model <chr>	AIC <dbl>
Full Model	784.55
polynomial Model	712.80
Interaction Model	565.02
Reduce polynomial	258.24

- The Reduced Polynomial model achieves the lowes. AIC (258.24), indicating the best trade-off between accuracy and simplicity.
- It significantly outperforms all other models, suggesting it is the most efficient and parsimonious choice.
- This supports selecting it as the final model for prediction and interpretation.

```
Training MSE:  0.2216361
Training RMSE:  0.4707824
Testing MSE:  0.1285106
Testing RMSE:  0.3584837
The model does not show signs of overfitt
```

In our CrossTrain70 evaluation, the training MSE is 0.2216 (RMSE = 0.4708), while the testing MSE is 0.1285 (RMSE = 0.3585), indicating good generalization and no signs of overfitting. Among the four model variations, the Full Model had an RMSE of 1.678, the Polynomial Model 1.361, the Interaction Model 0.950, and the Reduced Polynomial Model performed best with an RMSE of 0.441, offering the most accurate predictions.

```
Full Model   Polynomial Model  Interaction Model Reduced Polynomi
 1.6777162         1.3607698          0.9499015         0.4411685
```

- Sales=2.713+0.1996 · TV−0.002528 · TV2+0.04300 · Radio+0.00001633 · TV3−0.00000004973 · TV4+0.000000000 05702 · TV5+0.001025 · (TV×Radio)

Optimal Budget Allocation

Model optimization suggests that when the total budget is fixed, allocating all funds to TV advertising maximizes predicted sales. This indicates TV has a significantly higher marginal return than Radio. While the model recommends zero spending on Radio, real-world decisions should still consider market context. Overall, TV advertising proves most impactful in this dataset.

Limitations

- No info on market or economy factors
- No data on celebrity ads
- No data on how long people watched the ads.

Future Work

After collecting additional data, we will use the model to analyze unit pricing and estimate the optimal price point for maximizing profit.

THANK YOU

Lagrange Multiplier

$$a(TV, Radio) = 6.783 + 0.01925 \cdot TV + 0.02862 \cdot Radio + 0.001075 \cdot (TV \cdot Radio)$$

$$b(TV, Radio) = TV + Radio - \text{budget} = 0$$

$$c(TV, Radio) = a(TV, Radio) - \lambda \cdot b(TV, Radio)$$

$$\frac{\partial c}{\partial TV} = \frac{\partial a}{\partial TV} - \lambda \quad \frac{\partial c}{\partial Radio} = \frac{\partial a}{\partial Radio} - \lambda \quad \frac{\partial c}{\partial \lambda} = b(TV, Radio)$$

$$\frac{\partial c}{\partial TV} = \frac{\partial c}{\partial Radio}$$

$$0.01925 + 0.001075 \cdot Radio = 0.02862 + 0.001075 \cdot TV$$

$$Radio - TV = 8.72$$

1000 sales $TV = 800 \Rightarrow 8000$ $Radio = 200 \Rightarrow 300$
1000 sales $TV = 801 \Rightarrow 8000.5$ $Radio = 199 \Rightarrow 298.5$