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

Authors: David Teng, Hao Shen

Date: May 14, 2025

---

1. Introduction

Effective allocation of advertising budgets across different media platforms is a critical factor influencing product sales. In this study, we aim to analyze the impact of advertising expenditures on TV, Radio, and Newspaper channels on product sales using statistical modeling. The primary research questions guiding this analysis are: (1) Which advertising channels significantly contribute to increased sales? (2) What model best captures the relationship between advertising investments and sales performance? (3) How should advertisers allocate their budgets across media channels to maximize sales outcomes? Using multiple linear regression and advanced modeling techniques, we seek to answer these questions and provide actionable insights for marketing strategies.

2. Data Description

The dataset used in this analysis is the "Advertisement Sales Dataset," publicly available on Kaggle. It consists of 200 observations and five variables, collected in the United States during the early 2000s (2001-2006). The response variable is Sales (in thousands of units), and the predictor variables are TV, Radio, and Newspaper advertising expenditures (in thousands of dollars). Each row represents an independent observation from a regional market within that time frame.

Summary Statistics:

The following is a summary of the descriptive statistics for the current study. The main findings include that TV advertising budgets range from \$0.7k to \$296.4k while radio advertising budgets range from \$0k to \$49.6k, and newspaper advertising budgets range from \$0.3k to \$114k. As for sales, they range from 1.6k to 27k units across the three different categories.

Exploratory Data Analysis:

Scatterplots reveal a strong positive linear relationship between Sales and TV advertising expenditures, indicating that higher investments in TV campaigns are generally associated with

increased Sales. A moderate positive relationship is also observed between Sales and Radio advertising expenditures, albeit with more spread in the data. In contrast, the relationship between Newspaper advertising and Sales appears weaker and less consistent. See Figure 1 below.

Boxplots further illustrate these patterns, showing that Newspaper advertising budgets exhibit greater variability and include several high-end outliers, whereas TV and Radio budgets are more tightly and consistently distributed. These exploratory findings suggest that TV and Radio are likely to be more reliable predictors of Sales in comparison to Newspaper advertising, motivating their prioritization in subsequent modeling. See Figure 2 below.

Figure 1: Pairwise Scatterplot of Sales and Advertising Channels

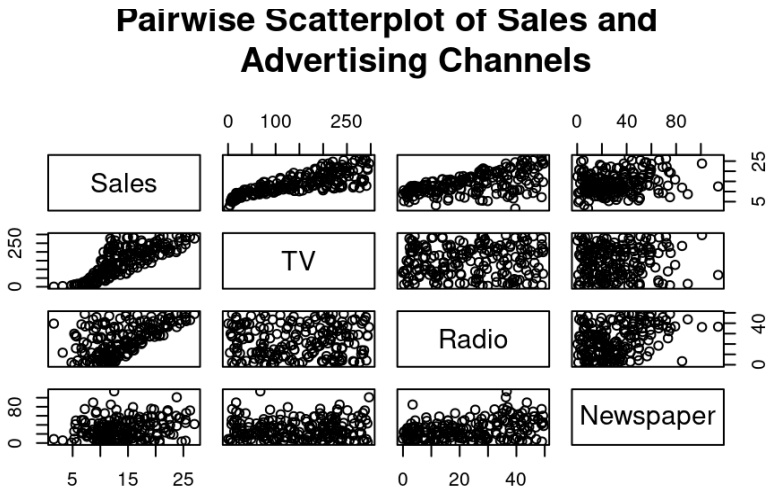
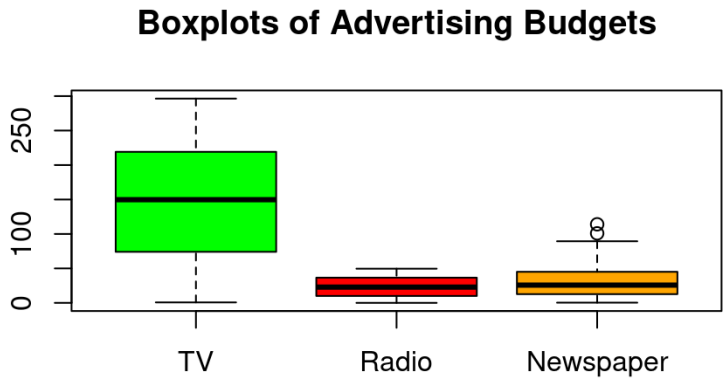


Figure 2: Boxplots of Advertising Budgets



3. Methods and Results

3.1. Multiple Linear Regression

A multiple linear regression model was fitted with Sales as the response and TV, Radio, and Newspaper as predictors. Both TV and Radio had significant positive coefficients ($p < 0.001$), while the coefficient for Newspaper did not show statistical significance ($p = 0.822$). The model achieved an Adjusted R^2 of 0.894, indicating that about 89.4% of the variance in Sales was explained.

3.2. Interaction Model

This multiple linear regression model fits well (Adjusted R^2 = 96.54%, $p <$ 2.2e-16), indicating that the main effects of TV, Radio, and Newspaper, along with their interaction terms, play a significant role in explaining the sales data.

In the model, TV advertising expenditure significantly positively affects sales. Additionally, the interaction effect between TV and Radio is also significant, suggesting a synergistic effect when both are used together. In contrast, Newspaper (print advertising) and its interaction terms are not significant, indicating that Newspaper advertising has a relatively minor impact on sales in this dataset.

Therefore, when simplifying the model, it may be reasonable to remove the non-significant Newspaper-related terms to reduce model complexity and mitigate the risk of overfitting

3.3. Polynomial Model

Polynomial regression analysis shows a strong nonlinear relationship between TV advertising and sales, with significant higher-order TV terms (TV^2 to TV^5) indicating diminishing returns. Radio has a significant positive linear effect, while its quadratic term is not significant. The TV × Radio interaction is highly significant ($p <$ 2e-16), suggesting a synergistic effect when both are used together. Newspaper advertising has no significant impact. The model fits the data extremely well, with an Adjusted R^2 of 0.9955, explaining about 99.6% of the variance in sales.

3.4. Reduced Polynomial Model

The reduced polynomial regression model, including TV, Radio, and their interaction (TV × Radio), achieves an excellent fit with an Adjusted R^2 of 0.9959 and a low residual standard error of 0.3418. TV advertising shows strong nonlinear effects with all higher-order terms (TV^2 to TV^5) being significant, indicating diminishing returns. Radio has a significant positive linear effect. Most notably, the TV × Radio interaction is highly significant ($p <$ 2e-16), revealing a synergistic effect when both channels are used together. This model offers high accuracy with fewer terms, making it both effective and efficient.

3.5. AIC test

The Akaike Information Criterion (AIC) values show that the reduced polynomial model has the lowest AIC (111.01), indicating the best balance between model fit and complexity. The polynomial model follows with AIC = 123.15, also showing strong performance. The interaction model (AIC = 565.02) and full model (AIC = 672.38) have significantly higher AIC values, suggesting that their added complexity does not sufficiently improve fit. Therefore, the reduced polynomial model is the most optimal based on AIC.

3.6. Cross-Validation

The cross-validation results for the reduced polynomial model indicate that the model performs well both during training and testing. The Training Mean Squared Error (MSE) is 0.2216, while the Testing MSE is 0.1285. The relatively low discrepancy between the two suggests that the model generalizes well to unseen data and does not suffer from overfitting.

3.7. Model Assumption Verification

To validate the assumptions of the reduced regression model, we first calculated the standardized residuals and identified potential outliers as those with absolute values greater than 3. These outliers were removed to improve the robustness and explanatory power of the model. The standardized residuals versus fitted plot showed that most data points were concentrated within the [-2, 2] range, indicating that the model satisfies the assumption of constant variance. See Figure 3 below.

Furthermore, we conducted a Shapiro-Wilk normality test on the residuals after outlier removal. The test yielded W = 0.98945 and a $p = 0.2006$. We fail to reject the null hypothesis that the residuals are normally distributed. Therefore, after excluding outliers, the model satisfies both the normality and constant variance assumptions of linear regression, demonstrating strong statistical validity. See Figure 4 and Figure 5 below.

Figure 3: Residual vs Fitted

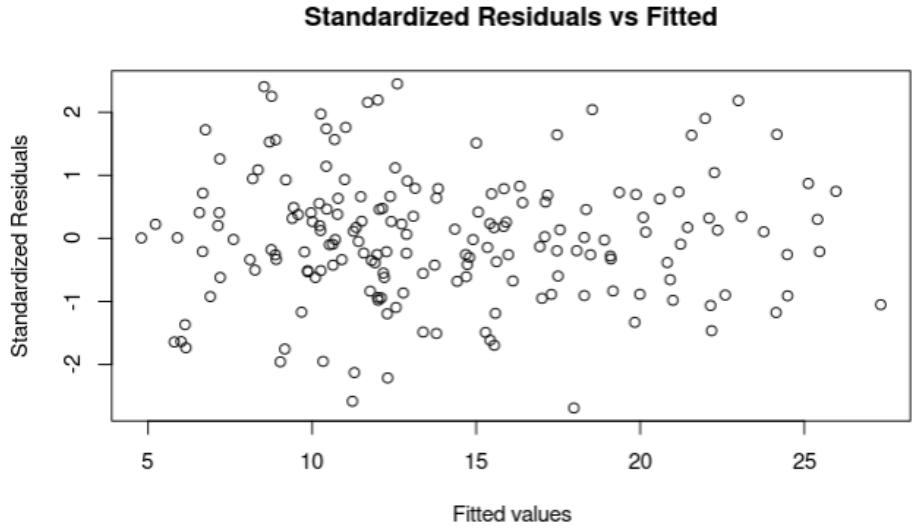


Figure 4: Q-Q plot

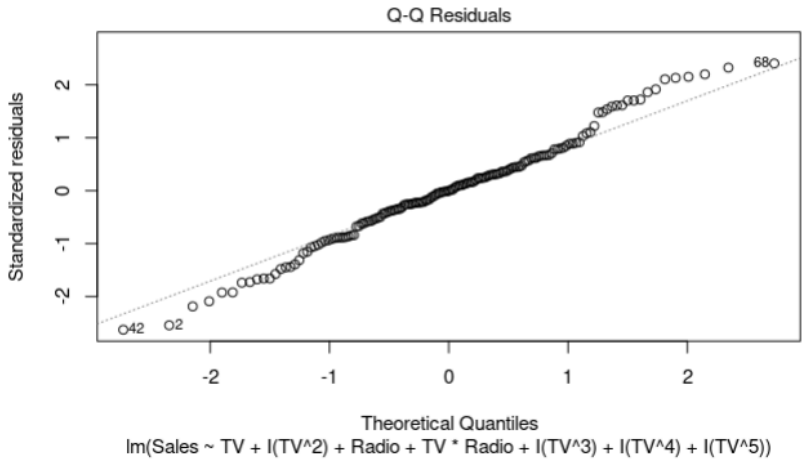


Figure 5: Reduced Polynomial Model

$$\text{Sales} = 2.713 + 0.1996 \cdot \text{TV} - 0.002528 \cdot \text{TV}^2 + 0.04300 \cdot \text{Radio} + 0.00001633 \cdot \text{TV}^3 - 0.00000004973 \cdot \text{TV}^4 + 0.00000000005702 \cdot \text{TV}^5 + 0.001025 \cdot (\text{TV} \times \text{Radio})$$

4. Conclusion

In conclusion, this study demonstrates that TV and Radio advertising are the primary drivers of sales, with a strong synergistic effect when used together. Polynomial modeling reveals nonlinear diminishing returns for TV, while Radio shows a primarily linear impact.

The reduced polynomial model offers the best trade-off between accuracy and simplicity (Adjusted R^2 = 0.9956, AIC = 111.01) and performs well in cross-validation, confirming its robustness and lack of overfitting.

Based on model optimization, the most effective budget allocation strategy is to invest entirely in TV advertising, as it yields the highest marginal return. While real-world decisions should consider broader factors, TV emerges as the most influential channel in this dataset.

5. Code Appendix

All R scripts used in the data analysis and model building are publicly available at the following GitHub repository:
https://github.com/David-W-Teng/STAT-632-Linear-and-Logistic-Regression

---

End of Paper