Pepsi Cola Project

James

2025-02-01

Introduction

The original dataset contained data for both Coca-Cola and Pepsi-Cola in a single sheet. To enhance clarity and facilitate analysis, the data was split into two separate sheets: 'brand1' for Coca-Cola and 'brand2' for Pepsi-Cola. Both sheets retain the same structure, ensuring consistency in analysis.

Question 1

Calculate descriptive (summary) statistics about the sales and use of marketing mix variables for both brands. Who is the market leader? How do the brands compare in terms of pricing, promotion, assortment? How do the brands compare in terms of allocation of spending between the four advertising instruments? Report the statistics and discuss your learnings and insights.

```
# Load necessary libraries
library(readxl)  # For reading Excel files
library(dplyr)   # For data manipulation
library(knitr)   # For displaying tables

# Read data from the Excel file
brand1 <- read_excel("MA_assignment_data.xls", sheet = "brand1")
brand2 <- read_excel("MA_assignment_data.xls", sheet = "brand2")

# Summary statistics for Coca-Cola (brand1) excluding 'week' column
summary_brand1 <- brand1 %>%
  select(-week) %>%  # Exclude the 'week' column
  summarise(across(where(is.numeric), list(
    Min = ~min(. , na.rm = TRUE),
    Max = ~max(. , na.rm = TRUE),
    Mean = ~mean(. , na.rm = TRUE),
    Median = ~median(. , na.rm = TRUE),
    SD = ~sd(. , na.rm = TRUE)
  )))

# Transpose the summary table for Coca-Cola
summary_brand1_transposed <- t(summary_brand1)

# Summary statistics for Pepsi-Cola (brand2) excluding 'week' column
summary_brand2 <- brand2 %>%
  select(-week) %>%  # Exclude the 'week' column
  summarise(across(where(is.numeric), list(
    Min = ~min(. , na.rm = TRUE),
```

```
    Max = ~max(. , na.rm = TRUE),
    Mean = ~mean(. , na.rm = TRUE),
    Median = ~median(. , na.rm = TRUE),
    SD = ~sd(. , na.rm = TRUE)
  )))

# Transpose the summary table for Pepsi-Cola
summary_brand2_transposed <- t(summary_brand2)

# Display results in Markdown format
cat("\n### Summary Statistics for Coca-Cola (Brand 1):\n")
```

```
##
## ### Summary Statistics for Coca-Cola (Brand 1):
```

```
kable(summary_brand1_transposed)
```

sales.brand1_Min	2.572088e+05
sales.brand1_Max	4.768800e+05
sales.brand1_Mean	3.082468e+05
sales.brand1_Median	2.986557e+05
sales.brand1_SD	3.885851e+04
feature.brand1_Min	0.000000e+00
feature.brand1_Max	6.584830e-01
feature.brand1_Mean	4.107210e-02
feature.brand1_Median	0.000000e+00
feature.brand1_SD	1.184135e-01
display.brand1_Min	8.018000e-04
display.brand1_Max	4.012522e-01
display.brand1_Mean	1.426173e-01
display.brand1_Median	1.315947e-01
display.brand1_SD	8.271070e-02
price.brand1_Min	1.273880e+00
price.brand1_Max	1.423048e+00
price.brand1_Mean	1.349152e+00
price.brand1_Median	1.350414e+00
price.brand1_SD	4.045660e-02
assortment.brand1_Min	3.830576e+01
assortment.brand1_Max	4.975457e+01
assortment.brand1_Mean	4.482845e+01
assortment.brand1_Median	4.480224e+01
assortment.brand1_SD	2.574797e+00
tv.brand1_Min	0.000000e+00
tv.brand1_Max	1.191093e+06
tv.brand1_Mean	3.781374e+05
tv.brand1_Median	3.339525e+05
tv.brand1_SD	2.925421e+05
digital.brand1_Min	0.000000e+00
digital.brand1_Max	5.002247e+04
digital.brand1_Mean	4.816267e+03
digital.brand1_Median	1.575005e+03

digital.brand1_SD	8.685430e+03
ooh.brand1_Min	0.000000e+00
ooh.brand1_Max	6.959014e+05
ooh.brand1_Mean	8.210566e+04
ooh.brand1_Median	8.883015e+03
ooh.brand1_SD	1.551237e+05
magazine.brand1_Min	0.000000e+00
magazine.brand1_Max	1.132490e+05
magazine.brand1_Mean	1.370076e+04
magazine.brand1_Median	0.000000e+00
magazine.brand1_SD	2.671565e+04

```
cat("\n### Summary Statistics for Pepsi-Cola (Brand 2):\n")
```

```
##
## ### Summary Statistics for Pepsi-Cola (Brand 2):
```

```
kable(summary_brand2_transposed)
```

sales.brand2_Min	3.336608e+04
sales.brand2_Max	8.431965e+04
sales.brand2_Mean	4.490184e+04
sales.brand2_Median	4.256765e+04
sales.brand2_SD	8.999532e+03
feature.brand2_Min	0.000000e+00
feature.brand2_Max	5.549139e-01
feature.brand2_Mean	4.374060e-02
feature.brand2_Median	0.000000e+00
feature.brand2_SD	1.472323e-01
display.brand2_Min	0.000000e+00
display.brand2_Max	6.503486e-01
display.brand2_Mean	4.505700e-02
display.brand2_Median	0.000000e+00
display.brand2_SD	1.256597e-01
price.brand2_Min	8.350984e-01
price.brand2_Max	1.126957e+00
price.brand2_Mean	1.044643e+00
price.brand2_Median	1.046526e+00
price.brand2_SD	5.682320e-02
assortment.brand2_Min	1.164443e+01
assortment.brand2_Max	1.440277e+01
assortment.brand2_Mean	1.295978e+01
assortment.brand2_Median	1.265810e+01
assortment.brand2_SD	8.157710e-01
tv.brand2_Min	0.000000e+00
tv.brand2_Max	4.354378e+05
tv.brand2_Mean	5.818956e+04
tv.brand2_Median	0.000000e+00
tv.brand2_SD	1.066801e+05
digital.brand2_Min	0.000000e+00

digital.brand2_Max	7.481380e+03
digital.brand2_Mean	2.288630e+02
digital.brand2_Median	0.000000e+00
digital.brand2_SD	9.415166e+02
ooh.brand2_Min	0.000000e+00
ooh.brand2_Max	2.006640e+05
ooh.brand2_Mean	7.687427e+03
ooh.brand2_Median	0.000000e+00
ooh.brand2_SD	3.466713e+04
magazine.brand2_Min	0.000000e+00
magazine.brand2_Max	8.600000e+04
magazine.brand2_Mean	7.818182e+02
magazine.brand2_Median	0.000000e+00
magazine.brand2_SD	8.199778e+03

Market Leadership:

Based on the summary statistics, **Coca-Cola** appears to be the market leader in terms of sales volume. The total weekly sales for Coca-Cola (brand1) have a **mean of 308,246 liters** and a **maximum of 476,880 liters**, while Pepsi-Cola (brand2) has a **mean of 44,901 liters** and a **maximum of 84,319 liters**. This significant difference in sales suggests Coca-Cola is the dominant brand in the market.

Pricing:

In terms of pricing, Coca-Cola has a slightly higher **average price** per liter at **€1.349** compared to Pepsi-Cola's **€1.045**. Coca-Cola's price range also spans from **€1.27** to **€1.42**, while Pepsi-Cola's price range is lower, ranging from **€0.83** to **€1.13**. This indicates that Coca-Cola is priced higher, potentially reflecting its premium market positioning.

Promotion:

For promotions, Coca-Cola has a significantly higher **average display percentage (14.26%)** compared to Pepsi-Cola's **4.51%**, which suggests that Coca-Cola has a stronger presence on the shelf in terms of visible displays. Additionally, Coca-Cola has a **higher average feature percentage (4.11%)** compared to Pepsi-Cola's **4.37%**. Despite this, Coca-Cola's stronger display presence suggests a more aggressive promotional strategy.

Assortment:

Coca-Cola also has a slightly larger **average assortment** (44.83 SKUs) compared to Pepsi-Cola's **12.96 SKUs**. This indicates that Coca-Cola offers a broader range of products, which could appeal to a wider variety of consumer preferences, giving it an advantage in terms of product availability and choice.

Advertising Spending Allocation:

When comparing advertising spend, Coca-Cola leads in most categories: - **TV advertising**: Coca-Cola spends an average of **€378,137** per week, significantly higher than Pepsi-Cola's **€58,190**. This suggests Coca-Cola is more aggressive in traditional media advertising. - **Digital advertising**: Coca-Cola also

has higher digital ad spending (**€4,816**) compared to Pepsi-Cola's **€288**. - **OOH advertising**: Coca-Cola spends **€82,106** on Out-Of-Home advertising, while Pepsi-Cola only spends **€7,687**. - **Magazine advertising**: Coca-Cola's magazine ad spending averages **€13,701**, compared to Pepsi-Cola's **€782**.

Overall, Coca-Cola allocates a higher proportion of its marketing budget across all four advertising instruments, with a particularly large focus on TV and OOH advertising.

Conclusion:

Coca-Cola is the clear market leader, outpacing Pepsi-Cola in sales volume, pricing, promotional activity, and advertising spend. The higher sales, larger assortment, more extensive promotional activity, and greater investment in advertising solidify Coca-Cola's dominance in the market. Pepsi-Cola, while competitive, appears to have a more restrained marketing strategy, focusing less on promotion and advertising.

Question 2

Estimate the following log-log regression model, which explains Pepsi's sales ("sales.brand2") as a function of Pepsi's own marketing mix variables. Note that a value of 1 is only added to variables that contain zero values (feature, display, and the 4 advertising instruments). (a) log(sales.brand2) = 0 + 1 log(feature.brand2+1) + 2 log(display.brand2+1) + 3 log(assortment.brand2) + 4 log(price.brand2) + 5 log(tv.brand2+1) + 6 log(digital.brand2+1) + 7 log(ooh.brand2+1) + 8 log(magazine.brand2+1) Report the estimates from equation (a) in a table. Based on this model, what do you conclude about the relation between Pepsi's marketing mix variables and its sales? How do you interpret the size of statistically significant estimates from equation (a)?

```
# Load necessary libraries
library(tidyverse)

# Transform data for log-log regression
brand2 <- brand2 %>%
  mutate(
    log_sales = log(sales.brand2),
    log_feature = ifelse(feature.brand2 == 0, log(feature.brand2 + 1), log(feature.brand2)),
    log_display = ifelse(display.brand2 == 0, log(display.brand2 + 1), log(display.brand2)),
    log_assortment = log(assortment.brand2),  # No +1 needed
    log_price = log(price.brand2),  # No +1 needed
    log_tv = ifelse(tv.brand2 == 0, log(tv.brand2 + 1), log(tv.brand2)),
    log_digital = ifelse(digital.brand2 == 0, log(digital.brand2 + 1), log(digital.brand2)),
    log_ooh = ifelse(ooh.brand2 == 0, log(ooh.brand2 + 1), log(ooh.brand2)),
    log_magazine = ifelse(magazine.brand2 == 0, log(magazine.brand2 + 1), log(magazine.brand2))
  )

# Estimate the log-log regression model
model <- lm(log_sales ~ log_feature + log_display + log_assortment + log_price +
              log_tv + log_digital + log_ooh + log_magazine, data = brand2)

# Display regression summary
summary(model)
```

```
##
## Call:
## lm(formula = log_sales ~ log_feature + log_display + log_assortment +
##     log_price + log_tv + log_digital + log_ooh + log_magazine,
```

```
##     data = brand2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.24270 -0.07678 -0.00239  0.05100  0.39468
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)    10.0394603  0.4802616  20.904  < 2e-16 ***
## log_feature    -0.7037532  0.0956596  -7.357 5.12e-11 ***
## log_display    -0.0055377  0.0069675  -0.795    0.429
## log_assortment  0.2455865  0.1852084   1.326    0.188
## log_price      -0.3369542  0.3022059  -1.115    0.268
## log_tv         -0.0014783  0.0022976  -0.643    0.521
## log_digital     0.0003863  0.0048891   0.079    0.937
## log_ooh         0.0053362  0.0034302   1.556    0.123
## log_magazine    0.0108600  0.0097461   1.114    0.268
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1073 on 101 degrees of freedom
## Multiple R-squared:  0.6412, Adjusted R-squared:  0.6128
## F-statistic: 22.56 on 8 and 101 DF,  p-value: < 2.2e-16
```

```r
# Load library for formatted tables
library(stargazer)
```

```
##
## Please cite as:

##  Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.

##  R package version 5.2.3. https://CRAN.R-project.org/package=stargazer
```

```r
# Generate a formatted regression table
stargazer(model, type = "text", title = "Log-Log Regression Results for Pepsi Sales")
```

```
##
## Log-Log Regression Results for Pepsi Sales
## ================================================
##                         Dependent variable:
##                     ---------------------------
##                               log_sales
## ------------------------------------------------
## log_feature                  -0.704***
##                               (0.096)
##
## log_display                   -0.006
##                               (0.007)
##
## log_assortment                 0.246
##                               (0.185)
```

```
##
## log_price                      -0.337
##                                (0.302)
##
## log_tv                         -0.001
##                                (0.002)
##
## log_digital                    0.0004
##                                (0.005)
##
## log_ooh                         0.005
##                                (0.003)
##
## log_magazine                    0.011
##                                (0.010)
##
## Constant                      10.039***
##                                (0.480)
##
## ------------------------------------------------
## Observations                     110
## R2                              0.641
## Adjusted R2                     0.613
## Residual Std. Error       0.107 (df = 101)
## F Statistic            22.562*** (df = 8; 101)
## ================================================
## Note:                *p<0.1; **p<0.05; ***p<0.01
```

Log-Log Regression Analysis of Pepsi Sales

1. Interpretation of the Results

Statistically Significant Variables ($p < 0.05$)

1. **Feature Advertising (`log_feature`)**
   - Estimate = **-0.704**, **p-value < 0.01 (highly significant)**
   - **Interpretation:** A 1% increase in **feature advertising** (promotions in-store) **decreases Pepsi's sales by ~0.70%**.
   - **Possible reason:** This could mean that promotional features do not directly drive sales, or they might be associated with price discounts that reduce total revenue.
2. **Intercept (constant term)**
   - Estimate = **10.039**, **p-value < 0.01**
   - This represents the baseline sales when all marketing variables are **at their minimum**.

Non-Significant Variables ($p > 0.05$)

- **Display, Assortment, Price, TV, Digital, OOH, and Magazine Advertising** are **not statistically significant** in this model, meaning they do not have a strong or consistent impact on sales.

---

2. Conclusion

- **Feature advertising** has a **negative and significant impact** on sales, suggesting that in-store promotions may not be effective or could be linked to price cuts.

- **Other marketing mix variables (display, price, TV, digital, OOH, and magazine advertising) do not show significant effects** on Pepsi's sales, at least in this dataset.
- The model explains **64.1%** of the variation in sales, which is fairly strong but suggests that **other factors (not included in this model) might also influence Pepsi's sales**.

---

3 Business Implications

- Pepsi might **re-evaluate feature promotions**, as they appear to **reduce sales** rather than increasing them.
- Since TV, digital, and other advertising types are **not significant**, Pepsi should consider **shifting its marketing strategy** or testing **new advertising channels**.

Would you like me to improve the model, such as testing interaction effects or checking for multicollinearity?