LEO-Wyndor Project

Xinjie Sun, Xinrui Yan, Yiwen Cao, Zelong Chen

Project Objectives

- Problem Definition
- Linear Regression
- Three different models
- Validation
- Comparison of solution
- Timeline

Problem Definition

Assumptions:

- Keep the range of expenditures same.
- All the variables are measured by thousand.
- Data is randomly split into 50-50 training and validation set for ease of drawing Q-Q plot.
- Drop all the outliers found in the Q-Q plot.
- Drop the newspaper column in the advertising dataset.

Linear Regression

- **We are using Linear Regression as a tool to help us predict the total sales (w_i) with the amount of advertising spent on TV (x_1) and Radio (x_2).**

- **Uncertainty is captured by the random error between the predicted and actual sales (ε_i)**

- **Randomly Split our data into 50-50 Training and Validation Set**

$$w_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon_i$$

Deterministic LP Formulation

- **Marketing Model**

$$
\begin{array}{llll}
Max & -0.5x_1 - 0.2x_2 + \mathbb{E}[\mathrm{Profit}(\tilde{\omega})] & & (1a) \\
s.t. & x_1 + x_2 & \le 200 & (1b) \\
 & x_1 - 0.2x_2 & \ge 0 & (1c) \\
 & -x_1 + 0.7x_2 & \ge 0 & (1d) \\
 & L_1 \le x_1 \le U_1, & L_2 \le x_2 \le U_2 & (1e)
\end{array}
$$

- **Production Model**

$$
\begin{array}{llll}
Profit(\omega) = Max & 3y_A + 5y_B & & (2a) \\
s.t. & y_A & \le 8 & (2b) \\
 & 2y_B & \le 24 & (2c) \\
 & 3y_A + 2y_B & \le 36 & (2d) \\
 & y_A + y_B & \le \omega & (2e) \\
 & y_A, y_B & \ge 0 & (2f)
\end{array}
$$

- **Integrated Model**

$$
\begin{array}{llll}
Max & -0.5x_1 - 0.2x_2 + 3y_A + 5y_B & & (3a) \\
s.t. & x_1 + x_2 & \le 200 & (3b) \\
 & x_1 - 0.2x_2 & \ge 0 & (3c) \\
 & -x_1 + 0.7x_2 & \ge 0 & (3d) \\
 & y_A & \le 8 & (3e) \\
 & 2y_B & \le 24 & (3f) \\
 & 3y_A + 2y_B & \le 36 & (3g) \\
 & -\beta_1 x_1 - \beta_2 x_2 + y_A + y_B & \le \beta_0 & (3h) \\
 & y_A, y_B & \ge 0 & (3i) \\
 & L_1 \le x_1 \le U_1, & L_2 \le x_2 \le U_2 & (3j)
\end{array}
$$

Stochastic Decomposition (SD) LP Formulation

- Based on the deterministic model, we add the error term in the SD model to represent the uncertainty of sales.
- Find the most suitable distribution from several distribution. (normal, lognormal, exponormal, exponential and gamma)
- Apply two sample chi-square test to both train set and validation set, proving that these two datasets follow the same normal distribution
- Calculate the mean and the standard deviation as 0.1073 and 1.3462

- Integrated Model

$$Max \; -0.5x_1 - 0.2x_2 + 3y_A + 5y_B \qquad (4a)$$

$$s.t. \quad x_1 + x_2 \leq 200 \qquad (4b)$$

$$x_1 - 0.2x_2 \geq 0 \qquad (4c)$$

$$-x_1 + 0.7x_2 \geq 0 \qquad (4d)$$

$$y_A \leq 8 \qquad (4e)$$

$$2y_B \leq 24 \qquad (4f)$$

$$3y_A + 2y_B \leq 36 \qquad (4g)$$

$$-\beta_1 x_1 - \beta_2 x_2 + y_A + y_B \leq \beta_0 + \varepsilon_{ti} \qquad (4h)$$

$$y_A, y_B \geq 0 \qquad (4i)$$

$$L_1 \leq x_1 \leq U_1, \quad L_2 \leq x_2 \leq U_2 \qquad (4j)$$

Sample Average Approximation Formulation

- Similar to the SD approach, SAA model also includes the error term which represents the uncertainty.
- The difference is that SAA calculates the expected profit. In order to avoid infinite sample size, we replace the expectation term by a finite representation of the average objectives.
- Assume the probabilities of all the outcomes are the same. (1/N)

- Integrated Model

$$Max - 0.5x_1 - 0.2x_2 + \frac{1}{N}\sum_{i=1}^{N} 3y_{Ai} + 5y_{Bi} \qquad (5a)$$

$$
\begin{aligned}
s.t \quad & x_1 + x_2 \leq 200 && (5b)\\
& x_1 - 0.2x_2 \geq 0 && (5c)\\
& -x_1 + 0.7x_2 \geq 0 && (5d)\\
& y_{Ai} \leq 8 \quad i = 1, \dots N && (5e)\\
& 2y_{Bi} \leq 24 \quad i = 1, \dots N && (5f)\\
& 3y_{Ai} + 2y_{Bi} \leq 36 \quad i = 1, \dots N && (5g)\\
& -\beta_1 x_1 - \beta_2 x_1 + y_{Ai} + y_{Bi} \leq \beta_0 + \varepsilon_{ti} \quad i = 1, \dots N && (5h)\\
& L_1 \leq x_1 \leq U_1, L_2 \leq x_2 \leq U_2 && (5i)\\
& y_{Ai}, y_{Bi} \geq 0 \quad i = 1, \dots N && (5j)
\end{aligned}
$$

Validation

- Calculate the uncertainty term with the validation data to generate possible outcomes (w_i) with the LR plus uncertainty and LP solution (x_1 x_2)
- For each w_i we can obtain the profit associated with each outcome by solving the Profit LP
- Using the profit and cost to find the average objective function value and standard deviation to create a 95% Confidence Interval

$$\varepsilon_{vi} = w_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i}$$

$$w_i := \beta_0 + \beta_1 \hat{x}_1 + \beta_2 \hat{x}_2 + \varepsilon_{vi}$$

$$\begin{aligned}
Profit(\omega) = Max \quad & 3y_A + 5y_B & & (2a)\\
s.t. \quad & y_A \le 8 & & (2b)\\
& 2y_B \le 24 & & (2c)\\
& 3y_A + 2y_B \le 36 & & (2d)\\
& y_A + y_B \le \omega & & (2e)\\
& y_A, y_B \ge 0 & & (2f)
\end{aligned}$$

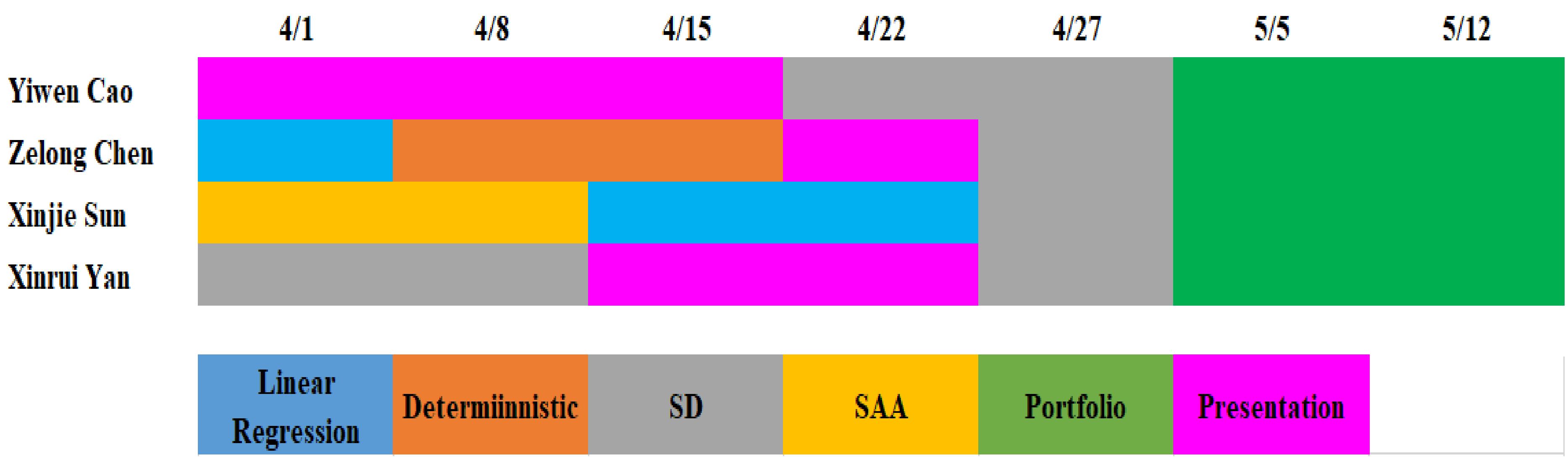
$$c_1\hat{x}_1 + c_2\hat{x}_2 + \frac{1}{N_v}\text{Profit}\sum_{i\in V}(\omega_i)$$

Comparison of different models

Methodology	x1	x2	MPO($)	MVSAE
Deterministic LP	9.92	49.60	49.636	49.403(±1.050)
SLP with SAA	9.92	49.60	49.628	49.723(±0.995)
SLP with SD	9.92	49.60	47.029	47.029(±0.259)

As can be seen from the table, the x1 and x2 value of these three models are almost the same, which is unusual. And the reasons are that the result of linear regression shows that radio ads contribute more to the sales and the cost of TV ads are greater than radio ads.

Timeline


4/1
4/8
4/15
4/22
4/27
5/5
5/12
Yiwen Cao
Zelong Chen
Xinjie Sun
Xinrui Yan
Linear Regression
Determiinnistic
SD
SAA
Portfolio
Presentation

Thank you for listening