Examples of Linear Regressions (Part II)

Le Wang

An Application: Advertising Budget and Sales

Suppose that we have a budget of $\$ 100,000$ for advertising. How shoul I, as a manager, decide where to spend such money?

1. On TV?
2. On Radio?
3. On Newspapaer?
4. On social media? How will the impact be even measured?

An Application: Advertising Budget and Sales

We need to have a well-defined objective before we even think about spending and conducting any empirical analysis.

One important **goal**: Sales of a Product

An Application: Advertising Budget and Sales

We are interested in predicting the sales for certain advertising budget

- Q1. Is there a relationship between advertising budget and sales?

Finally, we will introduce nonlinearity into our model

An Application: Advertising Budget and Sales

We are interested in predicting the sales for certain advertising budget

- Q1. Is there a relationship between advertising budget and sales?
- Q2. How strong is the relationship?

Finally, we will introduce nonlinearity into our model

An Application: Advertising Budget and Sales

We are interested in predicting the sales for certain advertising budget

- Q1. Is there a relationship between advertising budget and sales?
- Q2. How strong is the relationship?
- Q3. Which media contributes to sales? TV, Radio, or Newspaper

Finally, we will introduce nonlinearity into our model

An Application: Advertising Budget and Sales

We are interested in predicting the sales for certain advertising budget

- Q1. Is there a relationship between advertising budget and sales?
- Q2. How strong is the relationship?
- Q3. Which media contributes to sales? TV, Radio, or Newspaper
- Q4. Can we use this to predict future sales?

Finally, we will introduce nonlinearity into our model

An Application: Advertising Budget and Sales

We are interested in predicting the sales for certain advertising budget

- Q1. Is there a relationship between advertising budget and sales?
- Q2. How strong is the relationship?
- Q3. Which media contributes to sales? TV, Radio, or Newspaper
- Q4. Can we use this to predict future sales?

Finally, we will introduce nonlinearity into our model

- Q5. Is the relationship linear? (Nonlinearity)

An Application: Advertising Budget and Sales

We are interested in predicting the sales for certain advertising budget

- Q1. Is there a relationship between advertising budget and sales?
- Q2. How strong is the relationship?
- Q3. Which media contributes to sales? TV, Radio, or Newspaper
- Q4. Can we use this to predict future sales?

Finally, we will introduce nonlinearity into our model

- Q5. Is the relationship linear? (Nonlinearity)
- Q6. Is there synergy among the advertising media? (interaction effects)

Goal

Eventually we will use these estimates to decide on the budget allocations.

Data

The Advertising data set consists of the sales of that product in 200 different markets, along with advertising budgets for the product in each of those markets for three different media: TV, radio, and newspaper.

Data

```
##   X    TV Radio Newspaper Sales
## 1 1 230.1  37.8      69.2  22.1
## 2 2  44.5  39.3      45.1  10.4
## 3 3  17.2  45.9      69.3   9.3
## 4 4 151.5  41.3      58.5  18.5
## 5 5 180.8  10.8      58.4  12.9
## 6 6   8.7  48.9      75.0   7.2
```

Q1. Is there a relationship between advertising budget and sales?

```
##
## Call:
## lm(formula = Sales ~ TV, data = mydata)
##
## Coefficients:
## (Intercept)           TV
##     7.03259      0.04754
```

- Answer: Yes. There exists a positive relationship. The more you spend on TV advertising the more sales.

Q1. Is there a relationship between advertising budget and sales?

```
##
## Call:
## lm(formula = Sales ~ TV, data = mydata)
##
## Coefficients:
## (Intercept)           TV
##     7.03259      0.04754
```

- Answer: Yes. There exists a positive relationship. The more you spend on TV advertising the more sales.
- More specifically, an additional $1000 spent on TV advertising is associated with selling approximately 0.04754 additional units of the product.

Q2. How strong is the relationship?

If the evidence is weak, then one might argue that no money should be spent on advertising!

```
##
## Call:
## lm(formula = Sales ~ TV, data = mydata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.3860 -1.9545 -0.1913  2.0671  7.2124
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 7.032594   0.457843   15.36   <2e-16 ***
## TV          0.047537   0.002691   17.67   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
##
## Residual standard error: 3.259 on 198 degrees of freedom
## Multiple R-squared:  0.6119, Adjusted R-squared:  0.6099
## F-statistic: 312.1 on 1 and 198 DF,  p-value: < 2.2e-16
```

Q3. Which media contributes to sales? TV, Radio, or Newspaper

- Do all three media-TV, radio, and newspaper-contribute to sales,

Q3. Which media contributes to sales? TV, Radio, or Newspaper

- Do all three media-TV, radio, and newspaper-contribute to sales,
- Do just one or two of the media contribute?

To answer this question, we must find a way to separate out the individual (**marginal**) effects of each medium when we have spent money on all three media.

- $\text{sales} = \beta_0 + \beta_1 \cdot \text{TV Budget} + \beta_2 \cdot \text{Radio} + \beta_3 \cdot \text{Newspaper} + \epsilon$

To answer this question, we must find a way to separate out the individual (**marginal**) effects of each medium when we have spent money on all three media.

- sales $= \beta_0 + \beta_1 \cdot$ TV Budget $+ \beta_2 \cdot$ Radio $+ \beta_3 \cdot$ Newspaper $+ \epsilon$
- R code:

To answer this question, we must find a way to separate out the individual (**marginal**) effects of each medium when we have spent money on all three media.

- sales $= \beta_0 + \beta_1 \cdot$ TV Budget $+ \beta_2 \cdot$ Radio $+ \beta_3 \cdot$ Newspaper $+ \epsilon$
- R code:
- `lm(Sales ~ TV + Newspaper + Radio, data=mydata)`

```
##
## Call:
## lm(formula = Sales ~ TV + Newspaper + Radio, data = myda
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.8277 -0.8908  0.2418  1.1893  2.8292
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.938889   0.311908   9.422   <2e-16 ***
## TV           0.045765   0.001395  32.809   <2e-16 ***
## Newspaper   -0.001037   0.005871  -0.177     0.86
## Radio        0.188530   0.008611  21.893   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
##
## Residual standard error: 1.686 on 196 degrees of freedom
## Multiple R-squared:  0.8972, Adjusted R-squared:  0.8956
## F-statistic: 570.3 on 3 and 196 DF,  p-value: < 2.2e-16
```

Q3. Which media contributes to sales? TV, Radio, or Newspaper

Both TV and Radio contribute to sales, but not Newspaper. If anything, Newspaper budget even reduces sales.

Q4. Can we use this to predict future sales?

```
## Warning: package 'forecast' was built under R version 3.
```

Q4. Can we use this to predict future sales?

Suppose that according to our analysis, our company decides to spend a total of $100,000$ (thousands) dollars on advertising, $80,000$ (thousands) on radio, $20,000$ (thousands) on TV, and 0 on newspapers.

$$\text{forecast} = 2.9388894 + 0.0457646 \cdot 20000 - 0.0010375 \cdot 0 + 0.18853 \cdot 80000$$

- $\text{forecast} = 16000.63$

```
##   Point Forecast    Lo 80   Hi 80    Lo 95    Hi 95
## 1       16000.63 15115.57 16885.7 14643.23 17358.04
```

- Point forecast: 16000.63

```
##   Point Forecast    Lo 80   Hi 80    Lo 95    Hi 95
## 1       16000.63 15115.57 16885.7 14643.23 17358.04
```

- Point forecast: 16000.63
- Interval forecast at 80 percent confidence level: $[15115.57, 16885.7]$

```
##   Point Forecast    Lo 80   Hi 80    Lo 95    Hi 95
## 1       16000.63 15115.57 16885.7 14643.23 17358.04
```

- Point forecast: 16000.63
- Interval forecast at 80 percent confidence level: $[15115.57, 16885.7]$
- Interval forecast at 95 percent confidence level: $[14643.23, 17358.04]$

Marketing Example: Is there any synergy effect?

```
##
## Call:
## lm(formula = Sales ~ TV + Newspaper + Radio + Radio * TV
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.2929 -0.3983  0.1811  0.5957  1.5009
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6.728e+00  2.533e-01  26.561  < 2e-16 ***
## TV          1.907e-02  1.509e-03  12.633  < 2e-16 ***
## Newspaper   1.444e-03  3.295e-03   0.438  0.66169
## Radio       2.799e-02  9.141e-03   3.062  0.00251 **
## TV:Radio    1.087e-03  5.256e-05  20.686  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
##
```

Marketing Example: Is there any synergy effect?

```
##
## Call:
## lm(formula = Sales ~ TV + Newspaper + Radio + Radio * TV
##
## Coefficients:
## (Intercept)            TV     Newspaper         Radio
##    6.728412      0.019067      0.001444      0.027992
```

$$\text{Sales} = 6.728412 + 0.019067 \cdot \text{TV} + 0.001444 \cdot \text{Newspaper} + 0.027992 \cdot \text{Radio} + 0.001087 \cdot \text{TV} \cdot \text{Radio} + \epsilon$$

What does this model tell us?

$$\text{Sales} = 6.728412 + 0.019067 \cdot \text{TV} + 0.001444 \cdot \text{Newspaper} + 0.027992 \cdot \text{Radio} + 0.001087 \cdot \text{TV} \cdot \text{Radio} + \epsilon$$

- Scenario 1: TV $= 100{,}000$
  $\text{Sales} = 6.728412 + 0.019067 * 100000 = 1913.428$

What does this model tell us?

$$\text{Sales} = 6.728412 + 0.019067 \cdot \text{TV} + 0.001444 \cdot \text{Newspaper} + 0.027992 \cdot \text{Radio} + 0.001087 \cdot \text{TV} \cdot \text{Radio} + \epsilon$$

- Scenario 1: TV $= 100{,}000$
  $\text{Sales} = 6.728412 + 0.019067 * 100000 = 1913.428$
- Scenario 2: Radio $= 100{,}000$
  $\text{Sales} = 6.728412 + 0.027992 * 100000 = 2805.928$

What does this model tell us?

$$\text{Sales} = 6.728412 + 0.019067 \cdot \text{TV} + 0.001444 \cdot \text{Newspaper} + 0.027992 \cdot \text{Radio} + 0.001087 \cdot \text{TV} \cdot \text{Radio} + \epsilon$$

- Scenario 1: TV $= 100{,}000$
  $\text{Sales} = 6.728412 + 0.019067 * 100000 = 1913.428$
- Scenario 2: Radio $= 100{,}000$
  $\text{Sales} = 6.728412 + 0.027992 * 100000 = 2805.928$
- Scenario 3: TV $= 20{,}000$ and Radio $= 80{,}000$

What does this model tell us?

$$\text{Sales} = 6.728412 + 0.019067 \cdot \text{TV} + 0.001444 \cdot \text{Newspaper} + 0.027992 \cdot \text{Radio} + 0.001087 \cdot \text{TV} \cdot \text{Radio} + \epsilon$$

- Scenario 1: TV $= 100{,}000$
  $\text{Sales} = 6.728412 + 0.019067 * 100000 = 1913.428$
- Scenario 2: Radio $= 100{,}000$
  $\text{Sales} = 6.728412 + 0.027992 * 100000 = 2805.928$
- Scenario 3: TV $= 20{,}000$ and Radio $= 80{,}000$
-

$$\text{Sales} = 6.728412 + 0.019067 * 20000 + 0.027992 * 80000 + 0.001087 * 80000 * 20000 = 1,741,827$$

How to allocate my budget?

Turning back to my final goal: How to allocate my budget?

In the absence of synergy effects: What should you do?

In the presence of synergy effects: What should you do?

How to allocate my budget?

Optimal Weight

$$(w \cdot A) \cdot \beta_{TV} + (1 - w) \cdot A \cdot \beta_{Radio} + w \cdot A \cdot (1 - w) \cdot A\beta_{TV\,Radio}$$

We would like to choose w to maximize the impact on sales

How to allocate my budget?

Optimal Weight

$$
\begin{aligned}
0 &= \frac{\partial}{\partial w}[(w \cdot A) \cdot \beta_{TV} + (1 - w) \cdot A \cdot \beta_{Radio} \\
&\quad + w \cdot A \cdot (1 - w) \cdot A\beta_{TV\,Radio}] \\
&= A \cdot \beta_{TV} - A \cdot \beta_{Radio} + A^2\beta_{TV\,Radio} - 2w \cdot A^2 \cdot \beta_{TV\,Radio} \\
w &= \frac{A \cdot \beta_{TV} - A \cdot \beta_{Radio} + A^2\beta_{TV\,Radio}}{2w \cdot A^2 \cdot \beta_{TV\,Radio}} \\
&= \frac{100 * 0.0190668 - 100 * 0.0279917 + 100^2 * 0.0010873}{2 * 100^2 0.0010873} \\
&= 0.4589599
\end{aligned}
$$

How to allocate my budget?

Let's try different combinations:

$$\begin{aligned}\text{Sales} &= 6.7284119 + 0.4 * 100 * 0.0190668 + 0 * 100 * 0.0014442 \\ &\quad + 0.6 * 100 * 0.0279917 + 0.4 * 0.6 * 100^2 * 0.0010873 \\ &= 11.7801842\end{aligned}$$

$$\begin{aligned}\text{Sales} &= 6.7284119 + 0.6 * 100 * 0.0190668 + 0 * 100 * 0.0014442 \\ &\quad + 0.4 * 100 * 0.0279917 + 0.4 * 0.6 * 100^2 * 0.0010873 \\ &= 11.6016873\end{aligned}$$

$$\begin{aligned}\text{Sales} &= 6.7284119 + 0.4589599 * 100 * 0.0190668 + 0 * 100 * 0.0014442 \\ &\quad + 0.5410401 * 100 * 0.0279917 + 0.2483157 * 100^2 * 0.0010873 \\ &= 11.8179829\end{aligned}$$