

## 1. 데이터 summary

```
> summary(advertising_data, ncol=5)
```

X	TV	radio	newspaper	sales
Min. : 1.00	Min. : 0.70	Min. : 0.000	Min. : 0.30	Min. : 1.60
1st Qu.: 50.75	1st Qu.: 74.38	1st Qu.: 9.975	1st Qu.: 12.75	1st Qu.: 10.38
Median : 100.50	Median : 149.75	Median : 22.900	Median : 25.75	Median : 12.90
Mean : 100.50	Mean : 147.04	Mean : 23.264	Mean : 30.55	Mean : 14.02
3rd Qu.: 150.25	3rd Qu.: 218.82	3rd Qu.: 36.525	3rd Qu.: 45.10	3rd Qu.: 17.40
Max. : 200.00	Max. : 296.40	Max. : 49.600	Max. : 114.00	Max. : 27.00

해당 데이터셋은 각 항목(TV, radio, 신문)별 광고비와 그에 따른 판매량(sales)로 구성되어있다.

Summary 테이블을 봤을 때, 전반적으로 TV 광고비가 상대적으로 높은 값을 가져간다는 것을 알 수 있으며, radio와 신문 광고비의 비율은 유사해 보인다. 자세한 EDA는 생략하겠다.

## 2. 회귀 모델 output(coefficient)

```
> kable(coefficients)
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.9388894	0.3119082	9.4222884	0.0000000
TV	0.0457646	0.0013949	32.8086244	0.0000000
radio	0.1885300	0.0086112	21.8934961	0.0000000
newspaper	-0.0010375	0.0058710	-0.1767146	0.8599151

Response variable = Sales

predictor = TV + radio + newspaper

판매량에 대한 다중 선형 회귀모델의 coefficient, std error, t-value, p-value table.

P-value를 통한 유의성 검정 -> 비교분석을 통한 post-hoc을 거치겠다

p-value를 기준으로

1. 판매량에 유의미한 영향을 끼치는 변수. (P-value < 0.05)

- TV
- Radio

2. 판매량에 유의미한 영향을 끼치지 않는 변수 (p-value >= 0.05)

- Newspaper

00광고비가 1 단위(수치적으로) 증가할 때:

- TV: 판매량이 0.046 단위 증가
- Radio: 판매량이 0.188 단위 증가
- Newspaper: 판매량이 0.001단위만큼 감소

해당 모델에서 TV보다는 radio 광고비가 판매량 증가에 더 큰 영향을 미친다는 것을 알 수 있다

### 3. Correlation matrix

```
> kable(correlation_mat)
```

```
|          |          TV|          radio| newspaper|          sales|
|:-----:|:-----:|:-----:|:-----:|:-----:|
|TV        | 1.0000000| 0.0548087| 0.0566479| 0.7822244|
|radio     | 0.0548087| 1.0000000| 0.3541038| 0.5762226|
|newspaper | 0.0566479| 0.3541038| 1.0000000| 0.2282990|
|sales     | 0.7822244| 0.5762226| 0.2282990| 1.0000000|
```

- tv 광고비와 판매량은 0.782의 강한 양의 상관관계를 보임
- radio는 0.576으로 마찬가지로 양의 상관관계를 나타냄
- newspaper은 0.228로 상대적으로 약한 상관관계를 보임