

## 회귀분석 Output Table

```

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 ***
radio        0.188530   0.008611  21.893  <2e-16 ***
newspaper    -0.001037   0.005871  -0.177    0.86
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 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

```

sales 변수가 response variable이고 나머지 변수는 predictor variable일 때 회귀식은 다음과 같다.

$$\text{sales} = 2.9389 + 0.0458 \times \text{TV} + 0.1885 \times \text{radio} - 0.0010 \times \text{newspaper}$$

p-value를 살펴보면 각 독립 변수의 기여도를 확인할 수 있는데, TV와 radio의 p-value가 매우 작으므로 두 변수는 sales 예측에 유의미한 영향을 미치는 것을 알 수 있습니다. 그러나 newspaper의 p-value는 보통의 유의 수준 0.05보다 다소 크기 때문에 유의미한 변수라고 할 수 없습니다.

R-squared 값을 보면 0.8972이고 이는 주어진 독립 변수들이 종속 변수 변동의 약 89.72%를 설명함을 의미합니다.

Adjusted R-squared는 0.8956이므로 모델의 설명력이 높음을 의미합니다.

F-statistic 값은 570.2이고 p-value는 매우 작으므로 모델이 통계적으로 유의미함을 의미합니다.

잔차 표준 오차는 1.686이므로 실제 데이터 값과 예측값의 평균 간 오차가 작지 않음을 의미합니다.

## Correlation matrix

	TV	radio	newspaper	sales
TV	1.00000000	0.05480866	0.05664787	0.7822244
radio	0.05480866	1.00000000	0.35410375	0.5762226
newspaper	0.05664787	0.35410375	1.00000000	0.2282990
sales	0.78222442	0.57622257	0.22829903	1.0000000

상관 행렬을 분석하면, TV와 sales 간 상관 계수가 0.78이므로 강한 양의 상관 관계가 있고 radio와 sales 간 상관 계수는 0.57로 중간 수준의 양의 상관 관계를 보입니다. newspaper와 sales 간 상관 계수는 0.23으로 비교적 약한 양의 상관 관계를 보입니다. 그러므로 sales에 영향을 주는 변수는 TV와 radio임을 알 수 있습니다. 그리고 변수 간 상관 관계를 살펴

보면 다른 변수 간의 상관관계는 괜찮으나 newspaper와 radio가 약간의 상관관계를 가지고 있음에 유의해야 합니다. 그러나 대체적으로 다중공선성 문제를 일으킬 정도의 상관관계는 없습니다.