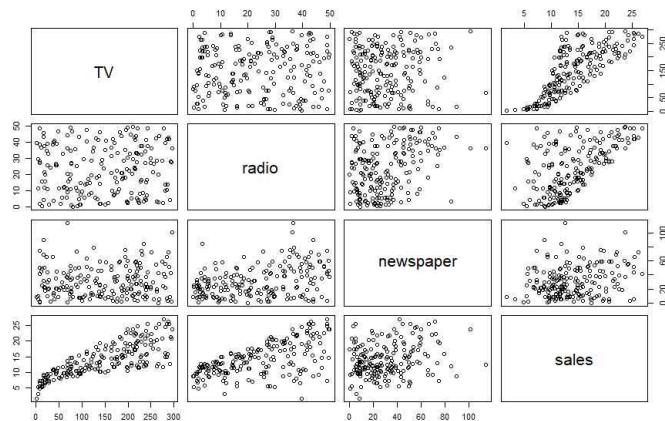


## 60152572 유정민 R통계분석 12주차 과제

문제 정의: 광고 카테고리 별 광고비 예산에 따른 예상 판매량 예측

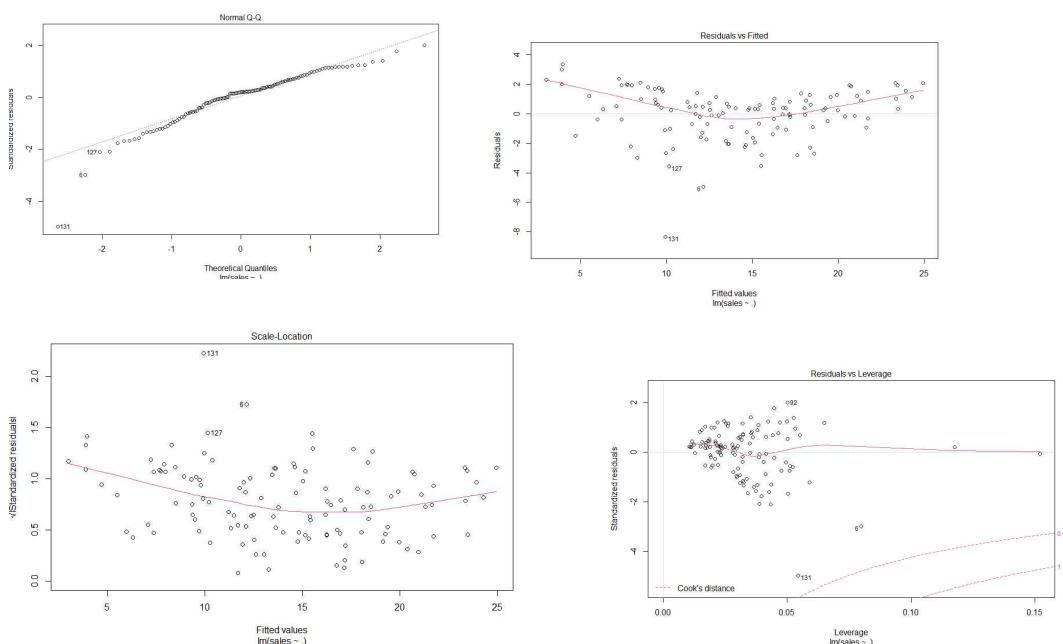
데이터 사전 확인: 문제 정의에 따라 sales~TV, sales~radio, sales~newspaper에 주목하여, 각각의 광고 수단이 sales와 대략적으로 연관이 있는지를 파악함



학습데이터: 원본데이터에서 무작위로 60% 추출

검정데이터: 원본데이터에서 학습데이터에 쓰이지 않은 나머지 40% 추출

sales를 종속변수로하고, 나머지 모두를 독립변수로 하는 다중선형회귀모델 생성 및 분석



Call:

```
lm(formula = sales ~ ., data = trainingData)
```

Residuals:

Min	1Q	Median	3Q	Max
-8.3310	-0.9121	0.3412	1.1234	3.3543

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.2753763	0.4211328	5.403	3.54e-07 ***
TV	0.0477543	0.0017575	27.171	< 2e-16 ***
radio	0.1923737	0.0115996	16.584	< 2e-16 ***
newspaper	0.0004856	0.0077866	0.062	0.95
---				

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.721 on 116 degrees of freedom

Multiple R-squared: 0.9006, Adjusted R-squared: 0.898

F-statistic: 350.3 on 3 and 116 DF, p-value: < 2.2e-16

--> Intercept와 TV, radio는 각각 sales를 설명하는 것에 통계적으로 유의미한 영향력이 있음.

--> newspaper는 p-value가 0.05보다 크므로 통계적으로 유의미한 영향력이 없음

--> 이 회귀모델은 90.06%의 데이터를 설명 가능함

--> F통계량의 p-value가 0.05보다 작으므로, 모델에 쓰인 독립변수들이 종속변수에 유의미한 영향력을 가짐을 의미.

예측모델생성: 학습데이터에 다중선형회귀모델을 적용한 예측모델을 생성함

예측데이터

->

1	3	4	9	11	12	14	16
20.568958	11.960351	17.483585	3.090533	6.559451	17.147125	8.396952	20.808469
20	30	32	34	35	36	40	45
13.916583	8.744617	11.032877	18.806526	7.118375	16.950398	20.431372	8.439037
46	51	56	57	59	61	62	63
14.980850	12.429837	21.306092	8.049785	21.902014	5.225368	22.994480	16.698018
65	68	69	71	72	74	76	81
16.783585	11.721915	18.907854	17.688675	10.285130	9.566506	11.532563	11.071007
84	85	86	89	91	95	101	104
14.119682	20.759390	15.073077	11.433248	9.635920	10.102708	13.747312	14.565920
105	107	108	110	112	115	116	118
20.251430	5.599765	6.661339	19.649336	21.139045	15.029599	12.620389	6.084887
119	121	125	126	128	129	130	137
15.415133	14.201101	19.484678	8.722133	6.109735	22.223502	7.450943	11.004974
139	141	142	143	144	149	152	153
9.321241	9.057155	18.372114	19.210399	8.383706	11.848475	9.693228	16.200819
157	160	163	165	166	168	170	171
15.152277	12.121089	14.766673	10.702690	14.168998	13.160720	17.894180	6.903558
172	179	182	183	184	187	196	197
14.174578	15.942947	13.761804	6.070117	24.316433	9.353997	4.818073	7.720392

실제데이터

->

```
[1] 22.1 9.3 18.5 4.8 8.6 17.4 9.7 22.4 14.6 10.5 11.9 17.4 9.5 12.8 21.5  
[16] 8.5 14.9 11.4 23.7 5.5 23.8 8.1 24.2 15.7 18.0 13.4 18.9 18.3 12.4 11.0  
[31] 8.7 11.8 13.6 21.7 15.2 12.9 11.2 11.5 11.7 14.7 20.7 7.2 8.7 19.8 21.8  
[46] 14.6 12.6 9.4 15.9 15.5 19.7 10.6 8.8 24.7 9.7 9.5 9.6 10.9 19.2 20.1  
[61] 10.4 10.9 11.6 16.6 15.3 12.9 14.9 11.9 11.9 12.2 15.0 8.4 14.5 11.8 12.2  
[76] 8.7 26.2 10.3 7.6 9.7
```

실제 데이터와 예측모델을 통한 결과값의 비교:

	actuals	predict
1	22.1	20.568958
3	9.3	11.960351
4	18.5	17.483585
9	4.8	3.090533
11	8.6	6.559451
12	17.4	17.147125

예측모델은 약95%의 정확도를 지님

	actuals	predict
actuals	1.0000000	0.9501138
predict	0.9501138	1.0000000