

-회귀분석의 목적: 주어진 데이터셋은 TV,라디오,신문에 관련된 수치형 데이터와 sales라고 명명된 판매수치를 가진 데이터셋이다. 200개의 케이스를 가지고 회귀분석을 진행한다. 회귀모델의 예측변수로 3개의 변수를 사용한다. 예측변수가 2개 이상이므로 해당 회귀분석은 다중회귀분석이라고 부른다. 해당 다중회귀 분석을 통해 TV매체에 지출된 광고비, 라디오 매체에 지출된 광고비, 신문 매체에 지출된 광고비가 판매량(sales)에 미치는 영향을 분석하여, 첫번째로 특정 예측변수 수준에서의 평균반응(mean response)를 추정할 수도, 그리고 두번째로 특정 예측변수 수준에서의 새로운 반응변수 관찰값(new observation)을 예측할 수 있다.

모수를 OLS방법을 통하여 구한 회귀계수에 대한 정보는 다음과 같다.

-회귀계수(각 독립변수가 종속변수에 미치는 영향)

-표준오차(회귀계수의 불확실성을 나타내는 지표로, 표준오차가 크다면 회귀계수의 신뢰도가 떨어짐을 의미하여 안정성 척도로 사용한다.)

-t통계량(회귀계수가 0인지 아닌지 그 유의미성을 가설검정을 통해 검정한 후 유의미하다고 판단되면 사용해야 하는데, 이 때 해당 모수가 0이라는 귀무가설을 대입한 t통계량을 사용하여 의사결정한다. 해당 t통계량과 p밸류를 잘 종합하여 계수를 0이라고 할 마땅한 이유가 없는지 검증하고 사용한다.)

OLS Regression Results						
=====						
Dep. Variable:	sales	R-squared:	0.897			
Model:	OLS	Adj. R-squared:	0.896			
Method:	Least Squares	F-statistic:	570.3			
Date:	Sat, 18 Jan 2025	Prob (F-statistic):	1.58e-96			
Time:	23:03:17	Log-Likelihood:	-386.18			
No. Observations:	200	AIC:	780.4			
Df Residuals:	196	BIC:	793.6			
Df Model:	3					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]
const	2.9389	0.312	9.422	0.000	2.324	3.554
TV	0.0458	0.001	32.809	0.000	0.043	0.049
radio	0.1885	0.009	21.893	0.000	0.172	0.206
newspaper	-0.0010	0.006	-0.177	0.860	-0.013	0.011
=====						
Omnibus:	60.414	Durbin-Watson:	2.084			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	151.241			
Skew:	-1.327	Prob(JB):	1.44e-33			
Kurtosis:	6.332	Cond. No.	454.			
=====						

해당 다중회귀모델은 표준오차를 통해 미루어 보았을 때 상수항의 불안정성이 타 변수들의 불안정성에 비해 조금 크고, $P>|t|$ 지표를 통해 해당 지표가 0.05보다 큰 신문 예측변수가 유의미하지 않다는 정보를 보여준다.

간단하게만 결론을 도출해보면, TV와 라디오는 그 광고의 효과가 t통계량을 통해 봤을 때 판매량에 확실한 영향을 미치지만, 신문 광고는 t밸류가 0에 가까울 뿐더러 $P>|t|$ 지표가 거의 1에 가깝기 때문에 독립변수로서 지니는 의미가 적다. 따라서 해당 기업의 광

고는 TV를 통해 하는 것이 가장 효과적이며, 신문은 축소해도 된다고 판단 가능하다.

해당 다중회귀분석에서의 correlation matrix는 multicollinearity를 확인하기 위해 체크한다. 다중공선성은 회피해야 할 성질이다. 독립변수들 간의 높은 상관관계가 있다면 회귀분석 결과가 왜곡되기 때문이다. 상관계수의 절댓값이 0.8이상이라면 독립변수간의 상관관계가 존재하여 차원축소 등의 조치를 취해줘야 한다.

	TV	radio	newspaper	sales
TV	1.000000	0.054809	0.056648	0.782224
radio	0.054809	1.000000	0.354104	0.576223
newspaper	0.056648	0.354104	1.000000	0.228299
sales	0.782224	0.576223	0.228299	1.000000

위 상관계수행렬을 통해 확인해볼 때, 직접적으로 0.8을 넘는 상관계수는 없다. 다중공선성을 걱정할 필요가 없다고 봐도 된다고 생각한다. 번외로 TV와 sales 변수끼리의 상관계수가 1에 가까운 것은 독립변수와 종속변수의 관계이며, TV광고가 판매량에 미치는 영향이 매우 큰 것을 유추할 수 있는 또 다른 지표라고 봐도 무방하다.