

기초 통계 과제 보고서

조석희

1. 회귀분석 결과 해석

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:	Fri, 17 Jan 2025	Prob (F-statistic):	1.58e-96			
Time:	00:40:47	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.				

sales 변수를 종속변수로, TV, radio, newspaper 변수를 독립변수로 설정하여 다중 회귀분석을 수행한 결과는 위와 같다.

먼저 R^2 값 0.897을 통해 해당 회귀 모델이 전체 변동에 대해 약 89%의 설명력을 갖는다고 해석할 수 있다. OLS 방법으로 잔차 제곱의 합의 최소를 구하는 과정에서 독립변수의 수가 증가할 때 R^2 값은 유지되거나 커지기 때문에 이를 보정하기 위한 Adjusted R^2 값이 0.896으로 소폭 감소했음을 확인할 수 있다.

F 값은 회귀 모델의 유의성을 평가하는 데 사용된다. F 값은 회귀선으로 설명 가능한 변량의 평균을 회귀선이 설명하지 못하는 변량의 평균으로 나눈 값이기 때문에 이 값이 클수록 모델의 설명력이 크고 독립 변수가 종속 변수에 유의미한 영향을 미친다는 것을 의미한다. 이러한 F 값의 p 값이 0.05 이하로 매우 작기 때문에 "모든 독립변수의 회귀 계수가 0"이라는 귀무가설을 기각하고 모델이 유의미함을 판단할 수 있다.

t 값은 회귀 분석에서 개별 독립변수가 종속변수에 미치는 영향을 측정하는 통계량이다. t 값에 대한 p 값이 0에 가까운 TV와 radio 변수는 sales 변수에 유의미한 영향을 미친다. 반면 해당 값이 0.860으로 매우 큰 값을 보이는 newspaper 변수는 sales 변수에 유의미한 영향을 미치지 않는다.

회귀 계수는 각 독립변수가 한 단위 변화할 때 종속변수에 미치는 영향을 나타낸다. 회귀 계수가 0.1885로 가장 큰 Radio 변수가 sales 변수에 가장 큰 영향을 미치며, TV 변수가 sales 변수에 미치는 영향은 두번째로 크다. 반면 newspaper의 회귀 계수는 0에 가까워 종속변수에 미치는 영향이 매우 작다.

2. 상관관계 행렬 해석

	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

각 변수간 상관계수를 행렬로 나타낸 사진은 위와 같다. TV 변수는 다른 독립변수들과의 상관 관계가 거의 없음을 알 수 있다. 그러나 sales 변수와는 강한 양의 상관 관계를 보이고 있는 점으로 미루어 보아 중요한 변수일 가능성이 있다는 것을 알 수 있다. 또한 sales와의 상관 계수에서 상대적으로 작은 값을 보이는 newspaper변수는 예측에 중요한 변수가 아니라는 것을 의미한다.