회귀분석 output table에서 coef는 coefficient를 의미하며, 독립 변수가 종속 변수에 얼마나 영향을 주는가를 나타냅니다. 즉, TV의 coef 값이 0.0458이라는 의미는 TV에 대한 광고비를 1단위 증가시키면 sales값(판매량)이 평균적으로 0.0458만큼 증가한다는 것을 뜻합니다. 반대로 newpaper는 coef 값이 –0.001이므로 신문에 대한 광고비를 1단위 증가시키면 평균적으로 판매량이 0.001만큼 감소한다는 뜻입니다.

std err는 표준오차를 뜻하며, coef의 불확실성을 의미합니다. t는 t-통계량을 의미하며, t-통계량의 절댓값이 클수록 보통 통계적으로 유의미함을 뜻하는데, 이것보단 뒤에 나올 p-value가 더 정확하게 측정할 수 있다. P>|t|에서 P는 p-value를 의미하는데, 이는 귀무가설이 참일 때, 현재 관측된 데이터보다 더 극단적인 값이 나올 확률을 의미합니다. 보통 0.05를 기각역으로 잡는데, 결과적으론, p-value가 0.05보다 작으면 해당 계수가 통계적으로 유의미함을 나타냅니다. 즉, const, TV, radio는 회귀계수에서 통계적으로 유의미하다고 볼 수 있지만, newspaper의 회귀계수는 유의마하다고 보기 힘듭니다.

회귀식으로 나타내면

sales = 2.9389 + 0.0458*TV + 0.1885*radio – 0.001*newspaper

로 나타낼 수 있습니다.

correlation matrix에서는 변수간의 상관계수를 알 수 있습니다. 상관계수는 –1에서 1사이 값을 가지며, 1에 가까울수록 양의 선형관계를 강하게 가지고, -1에 가까울수록 음의 선형관계를 강하게 가집니다. 0에 가깝다면, 두 변수 간에 선형관계가 있다고 보기 힘들다고 할 수 있습니다.

correlation matrix를 통해 TV와 sales와의 상관계수가 0.782224임을 알 수 있고, 두 변수 간에 강한 양의 상관관계를 가진다는 것을 알 수 있으며, newspaper와 sales는 0.228299임을 보면, 두 변수 간에 정말 약한 양의 선형관계를 가진다는 것을 알 수 있습니다. 물론, 독립변수 간의 상관계수를 보면 다중공선성도 확인할 수 있습니다. radio와 newspaper의 상관계수가 0.354104임을 알 수 있는데, 위에서 newspaper의 회귀계수가 유의미하지 않다는 점과 radio와 newspaper의 상관계수가 0.354104임을 고려하면, 회귀식에서 newspaper의 변수를 제거하는 것도 회귀식의 설명력을 높이는데, 영향을 줄 수 있습니다.

	coef	std err	t	P>|t|
const	2.9389	0.312	9.422	0.000
TV	0.0458	0.001	32.809	0.000
radio	0.1885	0.009	21.893	0.000
newspaper	-0.0010	0.006	-0.177	0.860

	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