

Chapter. 01

[Acquisition] 세일즈 데이터 분석 마케팅 믹스
모델링

분석 결과 해석 : 잘 된 분석인가?

FAST CAMPUS
ONLINE

직장인을 위한 파이썬 데이터 분석

강사. 최윤진

Chapter. 01

분석 결과 해석 : 잘 된 분석인가? 어떻게 개선 가능한가?

STEP 1

미디어별
광고비
EDA

STEP 2

분석
모델링
매체비로
세일즈
예측

STEP 3

분석 결과
해석
전용 방안

| Stats model의 선형회귀분석 결과를 해석해봅니다.
| 모든 의미를 파악하기 보다는 꼭 확인해야 할 것들 위주로 살펴봅니다.

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:	Thu, 12 Mar 2020	Prob (F-statistic):	1.58e-96
Time:	22:13:40	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]
Intercept	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.

- R-Square 0.897로 매우 높은 편
- P-value 0.05수준에서 유의한 변수는 TV, radio 입니다.
- newspaper는 유의하지 않는 것으로 나타났습니다. 즉, 신문광고가 매출액에 미치는 영향은 유의하지 않다고 할 수 있습니다.
- 회귀식은 다음과 같습니다. $\text{sales} = 2.9389 + 0.0458TV + 0.1885\text{radio} - 0.001*\text{newspaper}$

Stats model의 선형회귀분석 결과를 해석해봅니다.
모든 의미를 파악하기 보다는 꼭 확인해야 할 것들 위주로 살펴봅니다.

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:	Thu, 12 Mar 2020	Prob (F-statistic):	1.58e-96
Time:	22:17:04	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]
Intercept	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.

OLS Regression Results

Dep. Variable:	sales	R-squared:	0.897
Model:	OLS	Adj. R-squared:	0.896
Method:	Least Squares	F-statistic:	859.6
Date:	Thu, 12 Mar 2020	Prob (F-statistic):	4.83e-98
Time:	22:17:04	Log Likelihood:	-386.20
No. Observations:	200	AIC:	778.4
Df Residuals:	197	BIC:	788.3
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	2.9211	0.294	9.919	0.000	2.340	3.502
TV	0.0458	0.001	32.909	0.000	0.043	0.048
radio	0.1880	0.008	23.382	0.000	0.172	0.204

Omnibus:	60.022	Durbin-Watson:	2.081
Prob(Omnibus):	0.000	Jarque-Bera (JB):	148.679
Skew:	-1.323	Prob(JB):	5.19e-33
Kurtosis:	6.292	Cond. No.	425.1

OLS Regression Results

Dep. Variable:	sales	R-squared:	0.612			
Model:	OLS	Adj. R-squared:	0.610			
Method:	Least Squares	F-statistic:	312.1			
Date:	Thu, 12 Mar 2020	Prob (F-statistic):	1.47e-42			
Time:	22:17:04	Log-Likelihood:	-519.05			
No. Observations:	200	AIC:	1042.			
Df Residuals:	198	BIC:	1049.			
Df Model:	1					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

Intercept	7.0326	0.458	15.360	0.000	6.130	7.935
TV	0.0475	0.003	17.668	0.000	0.042	0.053

Omnibus:	0.531	Durbin-Watson:	1.935			
Prob(Omnibus):	0.767	Jarque-Bera (JB):	0.669			
Skew:	-0.089	Prob(JB):	0.716			
Kurtosis:	2.779	Cond. No.	338.			

- 유의하지 않은 변수 newspaper를 제거한 model2의 AIC, BIC가 가장 낮습니다.
- 여러개의 모델 중 선택을 할 때 AIC, BIC가 가장 낮은지 여부로 정하기도 합니다.
- 물론 AIC, BIC가 유일한 판단기준은 아니고 RMSE, CFI 등 다른 기준들과 함께 고려되어야 합니다.
- 결과에 따르면 p-value가 0.05 이상으로, 신문광고는 매출액 예측에 있어서 변수의 유무가 통계적으로 유의한 차이를 보이지 않습니다.
- 즉, 신문광고 마케팅과 매출액은 관련이 없다고 할 수 있습니다.

Stats model의 선형회귀분석 결과를 해석해봅니다.
모든 의미를 파악하기 보다는 꼭 확인해야 할 것들 위주로 살펴봅니다.

OLS Regression Results						
Dep. Variable:	sales	R-squared:	0.897			
Model:	OLS	Adj. R-squared:	0.896			
Method:	Least Squares	F-statistic:	570.2			
Date:	Thu, 12 Mar 2020	Prob (F-statistic):	1.58e-96			
Time:	22:38:21	Log-Likelihood:	-386.19			
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]
Intercept	2.8894	0.361	7.995	0.000	2.177	3.602
TV	0.0457	0.001	32.810	0.000	0.043	0.048
radio	0.1876	0.008	22.190	0.000	0.171	0.204
newspaper	0.0060	0.040	0.152	0.879	-0.072	0.084
Omnibus:	59.670	Durbin-Watson:	2.078			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	146.495			
Skew:	-1.320	Prob(JB):	1.55e-32			
Kurtosis:	6.258	Cond. No.	521.			

OLS Regression Results						
Dep. Variable:	sales	R-squared:	0.897			
Model:	OLS	Adj. R-squared:	0.896			
Method:	Least Squares	F-statistic:	570.4			
Date:	Thu, 12 Mar 2020	Prob (F-statistic):	1.53e-96			
Time:	22:38:21	Log-Likelihood:	-386.15			
No. Observations:	200	AIC:	780.3			
Df Residuals:	196	BIC:	793.5			
Df Model:	3					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	2.8057	0.495	5.671	0.000	1.830	3.782
TV	0.0457	0.001	32.822	0.000	0.043	0.048
radio	0.1874	0.008	22.510	0.000	0.171	0.204
log_newspaper	0.0414	0.142	0.291	0.772	-0.240	0.322
Omnibus:	59.391	Durbin-Watson:	2.074			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	144.867			
Skew:	-1.316	Prob(JB):	3.49e-32			
Kurtosis:	6.233	Cond. No.	731.			

- 신문에 대한 상관계수는 음에서 양으로 변했지만 여전히 P-value 0.05수준에서 유의하지 않습니다.
- newspaper는 유의하지 않는 것으로 나타났습니다. 즉, 신문광고가 매출액에 미치는 영향은 유의하지 않다고 할 수 있습니다.