3.4

1. 판매량과 광고 예산에 관계가 있는가?

텍스트, 스크린샷, 폰트, 라인이(가) 표시된 사진

자동 생성된 설명

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명



폰트, 텍스트, 라인, 화이트이(가) 표시된 사진

자동 생성된 설명

여기서 이걸 보려면 F-통계 값을 보면 되는데, p-value 가 낮게 나와서 광고랑 매출 사이에 어떤 관계가 있다는 것을 알 수 있음. # p 값 낮으면 낮을수록 변수가 더 깊은 의미 보유.

2. 얼마나 관계가 깊은가?

얼마나 각 변수들이 깊은 관계를 가지고 있는지 보려면,

RSE, R^2 값을 보면 되는데, 각 값들, RSE는 작고, R^2가 높은 값을 가지고 있기 때문에 관계가 깊다고 볼 수 있음.

#보통 R^2은 0.7 이상이면 모델이 적합하다는 것 의미.

3. 판매와 각 매체와의 연관은 어떻게 되는가?

이건 p-value 보면 되는데, TV, 라디오는 값 낮고, 신문은 높다.

TV랑 라디오는 매출과 관련 있고, 신문은 관계가 없음 의미.(표 3.4 봐볼 것)

4. 각 매체들과 매출 사이의 관계는 얼마나 큰가?

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

이거 보고 신뢰구간 계산하면 95% 신뢰구간 계산하면 0.043~0.049(TV)



0.172~0.206(radio), -0.013~0.011(신문)

이거 보면, TV랑 라디오는 범위값에 0 포함 안되는데, 신문은 범위 값에 0 들어감.

신문은 관계가 없다고 볼 수 있음.

VIF값 낮음 Collinearity 없다고 할 수 있음.(뭐 변수랑 별 관계 없다 이런 의미)



5. 얼마나 정확하게 예측 가능한가?

개별 변수마다의 예측을 보고자 한다면 ‘prediction interval’ 보고,

평균적인 예측을 보고자 한다면 ‘confidence interval’ 보면 됨.

Prediction interval은 입실론(오류)값을 갖고 있기 때문에 보통 구간이 더 넓음.

6. 관계는 선형인가?

라인, 국기, 디자인이(가) 표시된 사진

자동 생성된 설명

Residual plot 봐보면 되는데, 데이터 분포가 평면 판때기 위를 벗어나 있음.

그래서 비선형적이라고 볼 수 있음.

7. 광고에 시너지 효과가 있는가?

Additive 관계 🡺 각 변수들이 독립적이다.

그래서 각 변수들 잘 모아서 상호작용 항 넣어보고, p-값 또 봐보고, 값이 작으면 시너지 효과가 있다고 볼 수 있음.

결국 상호작용 항 잘 만들어서 넣으면, 모델이 좋아질 수 있다고 말함.

3.5

Linear regression은 예측 어렵지 않게 할 수 있는데, 실제 값이 예측 값과 쫌 멀리 떨어져 있다 하면 parameteric 방식을 적용해야 함.

Non- parameteric 방식인 KNN Regression은 Linear regression 보다 더 유연.

폰트, 텍스트, 화이트, 번호이(가) 표시된 사진

자동 생성된 설명 KNN 식.

스크린샷이(가) 표시된 사진

자동 생성된 설명

편향 분산 trade off 적으로 접근해보면,

K 값이 크면 예측을 부드럽게는 한다. But 편향값도 커짐. # K값 주위의 몇 개 데이터 보는지 생각.

그래서 데이터가 일차원 적인 형태로 단순하면 Linear regression으로.

라인, 도표, 그래프, 경사이(가) 표시된 사진

자동 생성된 설명

K값 커지면 편향값 증가, 분산값은 감소.

라인, 도표, 그래프, 경사이(가) 표시된 사진

자동 생성된 설명

실제 MSE값 보니 K값 높아질수록, 오차가 줄어듦 표현(1/k) 꼴이라서.

도표, 라인, 그래프, 평면도이(가) 표시된 사진

자동 생성된 설명

다차원에서도 같이 설명.

KNN Regression 설명인데, 다차원 일수록 MSE값 점점 높아지는게 보임.

어쩌면 이런 다차원에서는 Linear Regression 처럼 더 단순한게 좋을 수도 있다고 하는거 같은데,

이것도 잘 보고 해야한다 함.