**< 다중선형회귀분석 회귀계수 구하기 >**

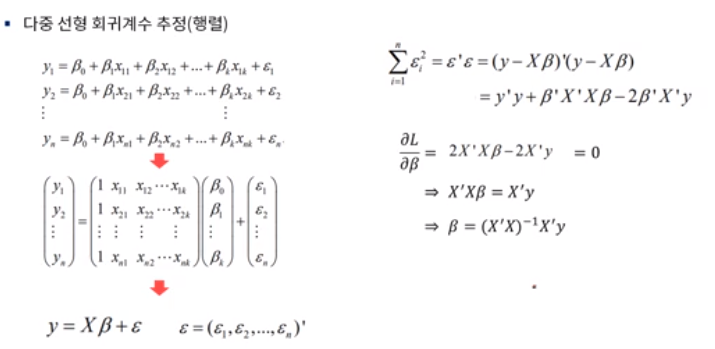
단순선형회귀분석 : 변수가 1개인 경우

Y\_hat = beta0 + beta1\*x

다중선형회귀분석 : 변수가 여러 개인 경우

Y\_hat = beta0+beta1\*x1+beta2\*x2 + …. + beta(p)\*x(p)

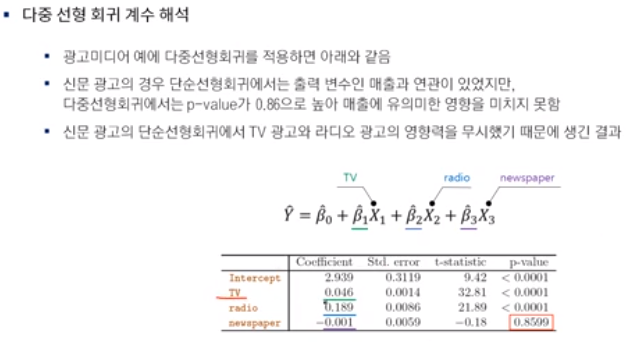
다중 선형 회귀계수 추정 : 회귀계수를 추정하는 것은 SSE를 최소화하는 방향으로 계산



Beta식을 결과로 보자 : 역행렬이 안구해지면 beta값이 의미가 없다

역행렬이 구해지지 않는 조건을 알아야한다.

**< 회귀계수와 모델의 검증 >**



다른 변수가 고정되어있을 때 TV가 1단위 증가할 때 매출액이 0.046단위 증가한다 . 그때 p-value가 매우 낮기 때문에 유의미한 변수로 볼 수 있다.

단순선형회귀분석에서는 하나의 변수에 대해 가설을 검증하면 모델도 검증이 되지만 다중선형회귀분석에서는 다른 변수들에대해서 검증이 되어야 모델을 검증할 수 있다.

다중 선형 회귀 모델 검정

* 귀무가설 : B1=B2=B3 … Bp=0
* 대립가설 : 하나의 회귀계수라도 0이 아니다. ( 매우 귀무가설을 기각하기 쉽다 ) 즉 설명력이있는 변수가 존재한다.
* F검정을 통해서 검정 F-statistics(MSR/MSE) 커질수록 p-value가 작아진다

??????????????????

변수가 추가되면 추가 될수록 귀무가설을 기각하기 쉽다 변수가 추가되면 MSR부분은 커지고 MSE부분은 작아진다 -> 따라서 F통계량은 커지게되고 p-value는 작아진다.

R-squared = SSR/SST 도 마찬가지로 커진다.

결국 변수가 매우많을때는 F통계량과 R^2를 마냥 믿을 수는 없다.

변수가 많아지면 자연스럽게 다중공선성 현상이 발생한다.

**< 다중공선성 >**

ㄷ