로짓모형(Logit model)과 토빗모형(Tobit model)

# 1. 로짓 모형 (Logit Model)

## 이론

로짓 모형은 종속변수가 이진(binary)인 경우, 즉 두 가지 가능한 결과(예: 성공/실패, 0/1)를 갖는 경우에 사용됩니다. 이 모형은 로지스틱 회귀(logistic regression)로도 불리며, 독립변수들이 종속변수의 발생 확률에 미치는 영향을 추정합니다. 이 모형의 주요 특징은 종속변수의 값이 확률로 해석될 수 있다는 점입니다.

## 수식

로짓 모형의 수식은 다음과 같습니다.  
- 로짓 함수: p(X) = 1 / (1 + e^-(β\_0 + β\_1 X\_1 + β\_2 X\_2 + ... + β\_k X\_k))  
 - 여기서 p(X)는 사건이 발생할 확률, X\_1, X\_2, ..., X\_k는 독립변수, β\_0, β\_1, ..., β\_k는 회귀계수입니다.  
- 이를 선형 회귀식으로 표현하면:  
 - logit(p) = log(p(X) / (1 - p(X))) = β\_0 + β\_1 X\_1 + β\_2 X\_2 + ... + β\_k X\_k

## 실제 사례

예를 들어, 소비자가 특정 제품을 구매할지 여부를 분석할 때 로짓 모형을 사용할 수 있습니다.  
1. 종속변수: 소비자가 제품을 구매했는지 여부(1 = 구매, 0 = 비구매).  
2. 독립변수: 가격, 광고 지출, 제품의 평점 등.  
3. 로짓 모형을 사용해 각 독립변수가 구매 확률에 미치는 영향을 추정합니다.  
 - 예를 들어, 광고 지출이 10% 증가할 때 구매 확률이 5% 증가할 수 있습니다.  
4. 최종적으로 모형의 결과를 통해 특정 소비자 그룹의 구매 확률을 예측할 수 있습니다.

# 2. 토빗 모형 (Tobit Model)

## 이론

토빗 모형은 종속변수가 절단(censored)되거나 관측된 값이 일정 범위 내에만 존재하는 경우에 사용됩니다. 예를 들어, 소득 데이터가 특정 하한이나 상한에서 절단된 경우 토빗 모형을 사용해 이러한 데이터를 분석할 수 있습니다.

## 수식

토빗 모형의 수식은 다음과 같습니다.  
- 종속변수 Y\_i\*는 잠재적인 연속 변수입니다.  
 - Y\_i\* = β\_0 + β\_1 X\_1 + β\_2 X\_2 + ... + β\_k X\_k + ε\_i  
 - 여기서 Y\_i\*는 잠재적 변수, X\_1, X\_2, ..., X\_k는 독립변수, β\_0, β\_1, ..., β\_k는 회귀계수, ε\_i는 오차항입니다.  
- 관측된 종속변수 Y\_i는 다음과 같이 정의됩니다:  
 - Y\_i = Y\_i\* if Y\_i\* > 0  
 - Y\_i = 0 if Y\_i\* ≤ 0

## 실제 사례

예를 들어, 가계 소비 분석에서 가계의 특정 상품에 대한 지출 금액을 분석할 때 토빗 모형을 사용할 수 있습니다.  
1. 종속변수: 특정 상품에 대한 가계 지출(0 이상).  
2. 독립변수: 가구 소득, 가구 구성원 수, 연령 등.  
3. 많은 가구가 특정 상품에 대해 지출하지 않을 수 있으므로, 이 경우 종속변수는 0으로 관측됩니다.  
4. 토빗 모형을 사용해 잠재적인 지출 금액을 추정하고, 특정 변수들이 지출 여부와 금액에 미치는 영향을 분석할 수 있습니다.

# 3. 로짓 모형과 토빗 모형의 비교

- 로짓 모형은 이진 종속변수(0 또는 1)에 사용됩니다. 주로 사건 발생 여부를 예측하거나 분석할 때 사용됩니다.  
 - 사용 예: 제품 구매 여부, 투표 여부 등.  
- 토빗 모형은 종속변수가 절단된 경우에 사용됩니다. 데이터가 특정 범위 내에서만 관측되는 경우 이 모형을 통해 정확한 추정을 할 수 있습니다.  
 - 사용 예: 특정 금액 이하로 절단된 소득 데이터, 특정 조건에서만 관측되는 시간 데이터 등.