# 내생성 문제 (Endogeneity Problem) 자세한 설명

## 1. 내생성 문제의 정의

내생성 문제는 독립변수가 종속변수에 영향을 미치는 동시에, 종속변수도 독립변수에 영향을 미치는 상호작용이 존재하는 경우 발생합니다. 이는 모델이 변수들 간의 진정한 인과관계를 제대로 반영하지 못하게 합니다.

수식적으로 내생성을 설명하는 방식:

일반적인 회귀 분석 모형:
Y = β0 + β1X + ε

여기서 Y는 종속변수, X는 독립변수, ε는 오차항입니다.

내생성 문제를 가진 모형:
X = γ0 + γ1Y + ν

이 경우, X가 Y에 영향을 미치는 동시에, Y도 X에 영향을 미치게 됩니다. 즉, X와 Y 간에 상호작용이 발생합니다.

이런 상호작용 때문에 ε과 ν가 상관관계를 가지게 되어, 회귀 분석의 기본 가정인 오차항의 독립성이 깨지게 됩니다. 이로 인해 회귀 계수 β1의 추정치가 편향(bias)될 가능성이 높아집니다.

## 2. 내생성 문제의 사례

경제학에서의 사례:

* 소득과 교육: 소득이 높은 사람은 더 나은 교육을 받을 가능성이 높습니다. 동시에, 더 높은 교육 수준은 더 높은 소득을 가져올 수 있습니다. 이런 경우, 소득과 교육 간에는 내생성이 존재하게 됩니다.
* 광고와 매출: 광고는 매출을 증가시키는 중요한 요인입니다. 하지만 매출이 증가하면 광고 예산도 증가할 수 있습니다. 이 경우, 광고와 매출 간에도 내생성이 존재하게 됩니다.

사회과학에서의 사례:

* 범죄율과 경찰 배치: 범죄율이 높으면 경찰을 더 많이 배치하게 됩니다. 하지만 경찰 배치가 증가하면 범죄율이 낮아질 수 있습니다. 이 관계 또한 내생성 문제를 나타냅니다.

## 3. 내생성 문제의 해결 방법

내생성 문제를 해결하기 위해 여러 가지 방법이 존재합니다. 대표적인 방법들은 다음과 같습니다.

### 1) 도구변수법 (Instrumental Variables Method)

도구변수: 도구변수는 내생변수와 상관관계가 있으나, 종속변수와는 직접적인 상관관계가 없는 외생변수를 사용하여 내생성을 해결합니다.

예시: 교육과 소득 간의 내생성 문제를 해결하기 위해, 도구변수로 출생 연도나 부모의 교육 수준을 사용할 수 있습니다. 이 변수들은 소득에 직접적인 영향을 미치지 않지만, 교육 수준과는 상관관계가 있습니다.

수식: 도구변수를 사용한 2단계 최소자승법 (Two-Stage Least Squares, 2SLS) 모델은 다음과 같이 표현됩니다.

1단계:
X = π0 + π1Z + u

2단계:
Y = β0 + β1X̂ + ε

여기서 Z는 도구변수, X̂는 1단계에서 예측된 X의 값입니다.

### 2) 패널 데이터 모델 (Panel Data Model)

고정효과 모형 (Fixed Effects Model): 패널 데이터를 이용하여 시간에 따라 변하지 않는 개체의 특성을 통제함으로써 내생성을 해결합니다.

차이의 차이 모형 (Difference-in-Differences, DiD): 두 개의 시점에서 실험군과 통제군의 변화를 비교하여 내생성을 제거합니다.

### 3) 무작위 대조 시험 (Randomized Controlled Trials, RCT)

무작위 대조 시험: 무작위로 실험군과 통제군을 설정하여 내생성을 방지하는 방법입니다. 이는 실험 설계에서 흔히 사용되며, 내생성 문제를 근본적으로 제거할 수 있습니다.

### 4) 내생성을 인지한 회귀모형

기타 방법들: 상관관계 구조를 고려한 다변량 회귀모형, 일반화 모멘트 방법 (Generalized Method of Moments, GMM) 등을 사용할 수 있습니다.

## 4. 결론

내생성 문제는 회귀 분석에서 매우 중요한 문제로, 잘못 처리하면 결과 해석에 심각한 오류를 초래할 수 있습니다. 이를 해결하기 위해서는 도구변수법, 패널 데이터 분석, 무작위 대조 시험 등의 방법을 활용하여 모델을 개선하는 것이 필요합니다. 이를 통해 내생성 문제를 해결하고, 보다 정확하고 신뢰할 수 있는 연구 결과를 얻을 수 있습니다.