

## Blinder-Oaxaca 분해방법의 이해

Blinder-Oaxaca 분해의 목적은 두 집단의 평균 결과 차이를 설명 변수 수준의 집단 차이와 회귀 계수 크기의 차이로 분해하여 그 원인을 설명하는 것이다(Oaxaca 1973; Blinder 1973; Hlavac, 2022). 여기서, Blinder-Oaxaca 분해의 관심정량인 평균 결과 차이( $\Delta \hat{Y}$ )는 집단 A와 집단 B의 관측치에 대한 평균 결과의 차이를 말한다(수식 (1) 참조).

$$\Delta \hat{Y} = \hat{Y}_A - \hat{Y}_B \quad (1)$$

선형 회귀에서 한 집단(G)의 기대 결과는  $\hat{Y}_G = \bar{X}'_G \hat{\beta}_G$ 로 표현할 수 있는데, 설명 변수의 평균 값( $\bar{X}_G$ )과 추정 회귀 계수( $\hat{\beta}_G$ )의 곱이다. 이를 바탕으로 두 집단간의 기대결과의 차이는 수식 (2)와 같이 다시 표현할 수 있다.

$$\Delta \hat{Y} = \bar{X}'_A \hat{\beta}_A - \bar{X}'_B \hat{\beta}_B \quad (2)$$

위 수식에서 다시 삼중분해(수식 (3) 참조)와 이중분해(수식(4)로 나뉘어진다.

$$\Delta \hat{Y} = \underbrace{(\bar{X}_A - \bar{X}_B)' \hat{\beta}_B}_{\text{endowments}} + \underbrace{\bar{X}'_B (\hat{\beta}_A - \hat{\beta}_B)}_{\text{coefficients}} + \underbrace{(\bar{X}_A - \bar{X}_B)' (\hat{\beta}_A - \hat{\beta}_B)}_{\text{interaction}} \quad (3)$$

삼중분해는 집단간 평균차이는 집단간 설명변수의 평균 차이의 기여부분(endowments), 회귀계수의 집단 차이 부분(coefficients), 기여부분과 회귀계수부분의 상호작용(interaction)으로 분해한다.

$$\Delta \hat{Y} = \underbrace{(\bar{X}_A - \bar{X}_B)' \hat{\beta}_R}_{\text{explained}} + \underbrace{\bar{X}'_A (\hat{\beta}_A - \hat{\beta}_R)}_{\text{unexplained A}} + \underbrace{\bar{X}'_B (\hat{\beta}_B - \hat{\beta}_R)}_{\text{unexplained B}} \quad (4)$$

이중분해는 계수의 기준 벡터( $\hat{\beta}_R$ )을 추가함으로써, 설명변수의 집단간 차이에 의해 설명되어지는 부분(explained)과 설명변수의 집단간 차이에 의해 설명되어지지 않는 부분(unexplained)으로 분해한다. 여기서, 후자는 다시 A 집단의 설명변수 평균에 의해 설명되지 않는 영역(unexplained A)과 B 집단의 설명변수 평균에 의해 설명되지 않는 영역(unexplained B)로 다시 분해할 수 있다.