# VIF가 10이 넘을 때 데이터를 보정하는 방법

## 1. 변수 제거

설명: 공선성이 높은 변수를 모델에서 제거합니다.

방법: VIF 값이 가장 높은 변수를 하나씩 제거하면서 VIF 값을 확인합니다.

예시: 다중 회귀 분석에서 특정 변수가 다른 변수들과 높은 상관관계를 가지면 그 변수를 제거합니다.

## 2. 변수 결합

설명: 공선성이 높은 변수들을 결합하여 새로운 변수를 만듭니다.

방법: 주성분 분석(PCA)이나 차원 축소 기법을 사용하여 변수를 결합합니다.

예시: 여러 관련 변수를 하나의 주성분으로 결합하여 분석에 사용합니다.

## 3. 정규화

설명: 변수를 정규화하여 공선성을 줄입니다.

방법: Z-score 정규화, Min-Max 정규화 등을 사용합니다.

예시: 각 변수를 평균이 0, 분산이 1이 되도록 변환합니다.

## 4. 다중공선성 완화

설명: 변수 간의 상관관계를 줄이기 위해 다중공선성을 완화하는 기법을 사용합니다.

방법: Ridge 회귀 분석, Lasso 회귀 분석 등을 사용합니다.

예시: Ridge 회귀 분석은 회귀 계수에 페널티를 부여하여 공선성을 줄입니다.

## 5. 상관관계 분석

설명: 변수 간의 상관관계를 분석하여 공선성이 높은 변수를 파악하고 조정합니다.

방법: 상관계수 행렬을 확인하고 공선성이 높은 변수들을 조정합니다.

예시: 상관계수 행렬에서 높은 상관관계를 가지는 변수 쌍을 확인하여 변수를 제거하거나 결합합니다.

## 6. 데이터 수집

설명: 추가 데이터를 수집하여 공선성을 줄입니다.

방법: 기존 변수 외에 새로운 독립변수를 추가하여 데이터의 다양성을 높입니다.

예시: 기존 데이터에 새로운 특성을 추가하여 공선성을 완화합니다.

## 7. 단계적 회귀 분석

설명: 단계적 회귀 분석을 통해 변수 선택을 최적화합니다.

방법: 전진 선택법, 후진 제거법, 단계적 선택법 등을 사용합니다.

예시: 단계적 선택법을 사용하여 최적의 변수 조합을 찾습니다.