Grubbs 검정과 Dixon's Q Family 검정

# 1. Grubbs 검정

## 이론

Grubbs 검정은 단일 특이치를 식별하기 위한 통계적 방법입니다. 이 검정은 주어진 데이터 세트에서 가장 극단적인 값(최대 또는 최소 값)이 특이치인지 여부를 평가하는 데 사용됩니다. Grubbs 검정은 주로 정규 분포를 가정하며, 평균과 표준 편차를 사용하여 특이치 여부를 판단합니다.

## 수식

Grubbs 검정에서 사용되는 통계량은 다음과 같습니다:  
  
G = |X\_i - X̄| / S  
  
- X\_i: 특이치로 의심되는 관측값  
- X̄: 데이터의 평균  
- S: 데이터의 표준 편차  
  
Grubbs 통계량 G는 가장 극단적인 값과 평균 사이의 차이를 표준 편차로 나눈 값입니다. 이 값이 크면 해당 데이터 포인트가 특이치일 가능성이 높습니다.

## 검정 절차

1. 데이터 세트에서 가장 극단적인 값을 선택합니다(최대 또는 최소).  
2. Grubbs 통계량 G를 계산합니다.  
3. 사전 지정된 유의 수준(예: α = 0.05)에서 G 통계량을 위한 임계값과 비교합니다.  
- 만약 G가 임계값보다 크면, 해당 데이터 포인트는 특이치로 간주됩니다.

## 사례

예를 들어, 다음과 같은 데이터가 있다고 가정해 봅시다:  
  
[2, 3, 4, 5, 6, 100]  
  
여기서 100은 다른 값들과 비교하여 매우 큰 값이므로 특이치로 의심됩니다.  
  
1. 평균 X̄ = 20  
2. 표준 편차 S ≈ 36.76  
3. Grubbs 통계량 G = |100 - 20| / 36.76 ≈ 2.18  
  
유의 수준 0.05에서 임계값이 2.02일 때, G > 2.02이므로 100은 특이치로 간주될 수 있습니다.

# 2. Dixon's Q Family 검정

## 이론

Dixon's Q 검정은 소규모 데이터 세트에서 한 개 또는 두 개의 특이치를 탐지하는 데 사용됩니다. 이 검정은 데이터 포인트 사이의 간격을 비교하여 특이치를 식별합니다. 주로 정렬된 데이터에서 가장 작은 값이나 가장 큰 값이 특이치인지 평가하는 데 사용됩니다.

## 수식

Dixon's Q 통계량은 다음과 같이 정의됩니다:  
  
Q = (X\_n - X\_{n-1}) / (X\_n - X\_1) 또는 Q = (X\_2 - X\_1) / (X\_n - X\_1)  
  
- X\_1, X\_2, ..., X\_n: 정렬된 데이터 포인트  
- 첫 번째 식은 최대값이 특이치로 의심될 때 사용  
- 두 번째 식은 최소값이 특이치로 의심될 때 사용

## 검정 절차

1. 데이터 세트를 오름차순으로 정렬합니다.  
2. Dixon's Q 통계량을 계산합니다.  
3. 사전 지정된 유의 수준(예: α = 0.05)에서 Q 통계량의 임계값과 비교합니다.  
- 만약 Q가 임계값보다 크면, 해당 데이터 포인트는 특이치로 간주됩니다.

## 사례

다음과 같은 데이터 세트를 고려해 봅시다:  
  
[1.2, 1.3, 1.4, 1.5, 1.6, 3.0]  
  
여기서 3.0이 특이치로 의심됩니다. 이 데이터를 오름차순으로 정렬하면 이미 정렬된 상태입니다.  
  
1. Dixon's Q 통계량 계산:  
Q = (3.0 - 1.6) / (3.0 - 1.2) = 1.4 / 1.8 ≈ 0.778  
2. 유의 수준 0.05에서 임계값은 약 0.568 (데이터 크기에 따라 다름).  
- Q > 0.568이므로 3.0은 특이치로 간주됩니다.

# 비교 및 결론

Grubbs 검정:  
- 적용 가능 상황 : 주로 단일 특이치를 확인하기 위한 검정이며, 데이터가 정규 분포를 따르는 경우에 적합합니다.  
- 장점 : 계산이 상대적으로 간단하고, 잘 알려진 통계량을 사용합니다.  
- 단점 : 여러 특이치가 존재할 때에는 적합하지 않을 수 있습니다.  
  
Dixon's Q 검정:  
- 적용 가능 상황: 주로 작은 데이터 세트에서 한 개 또는 두 개의 특이치를 탐지하는 데 유용합니다.  
- 장점 : 소규모 데이터 세트에서 효과적이며, 정규성 가정이 필요하지 않습니다.  
- 단점 : 큰 데이터 세트에서는 효과가 떨어질 수 있습니다.  
  
이 두 검정은 각각의 데이터 특성과 분석 목적에 따라 적절히 선택하여 사용됩니다. Grubbs 검정은 일반적으로 더 큰 데이터 세트와 정규 분포 가정이 있는 경우에 적합하며, Dixon's Q 검정은 작은 데이터 세트에서 특히 유용합니다.