

Disparate Impact Analysis

Disparate Impact: A phenomenon where seemingly neutral policies or practices have a disproportionately negative effect on specific groups

4/5ths Rule

Disparate Impact Ratio

$$P(\hat{Y}=1|Group=B) / P(\hat{Y}=1|Group=A)$$

✓ ≥ 0.8 : Fair

✗ < 0.8 : Suspected Disparate Impact

Detection Methods

- Statistical Significance Testing (Chi-square test)
- Comparison of Performance Metrics by Group
- Intersectionality Analysis
- Tracking Changes Over Time
- Causal Analysis

⚠ Examples in Medical AI

- Lower diagnosis rates for specific races
- Limited treatment access for elderly patients
- Under-treatment of female patients
- Lower recommendation grades for low-income patients