

# Wilcoxon 부호-순위 검정 (Wilcoxon Signed-rank Test)

- **\*\*대응표본 T-검정(Paired t-test)\*\*의 비모수(non-parametric) 버전**
- **데이터의 분포에 대한 가정이 필요 없음**
  - 모수 검정인 T-검정은 데이터의 정규성(normality) 가정이 필요
  - 대신, 데이터의 실제 값 대신 그 값들의 **\*\*순위(rank)\*\***를 이용하여 가설을 검정
- 주로 동일한 대상에 대한 두 번의 측정값(e.g. 처치 전/후)의 **\*\*중앙값(median)\*\***에 차이가 있는지를 확인하는 데 사용

## 검정 원리

1. 각 대응 쌍(pair) 간의 차이(difference)를 계산합니다.
2. 차이가 0인 경우는 제외합니다.
3. 차이의 절대값에 대해 순위를 매깁니다. (가장 작은 절대값부터 1, 2, 3, ...)
4. 원래 차이값의 부호(+ 또는 -)를 각 순위에 다시 붙입니다.
5. 양수(+) 순위들의 합과 음수(-) 순위들의 합을 구합니다.
6. 이 두 합 중 더 작은 값을 검정통계량(W)으로 사용하고, 이를 기반으로 p-value를 계산하여 가설을 검정합니다.

## 가설 설정

- **귀무가설 ( $H_0$ ):** 두 대응표본의 중앙값(또는 분포)은 동일하다.
  - 차이의 분포는 0을 중심으로 대칭이다.
- **대립가설 ( $H_1$ ):** 두 대응표본의 중앙값(또는 분포)은 다르다.

## 단일표본 Wilcoxon 부호-순위 검정

대응표본뿐만 아니라, 단일 표본의 중앙값이 특정 기준값( $m_0$ )과 같은지를 검정하는 데에도 사용할 수 있습니다. 이는 **\*\*단일표본 T-검정(One-sample t-test)\*\***의 비모수 버전입니다.

- **귀무가설 ( $H_0$ ):** 표본의 중앙값은  $m_0$ 와 같다.
- **대립가설 ( $H_1$ ):** 표본의 중앙값은  $m_0$ 와 다르다.

## 적용 가능한 상황

- **대응표본 T-검정의 정규성 가정이 만족되지 않을 때:** 약물 투여 전/후의 혈압 변화, 특정 교육 프로그램 참가 전/후의 시험 성적 변화 등을 비교할 때, 그 차이값의 분포가 정규분포를 따르지 않는 경우에 사용됩니다.
- **데이터가 순서형(Ordinal) 척도일 때:** 데이터가 실제 값은 아니지만 순위 정보(e.g., 만족도: 매우 불만족, 불만족, 보통, 만족, 매우 만족)를 가질 때 사용할 수 있습니다.
- **표본 크기가 매우 작을 때:** 표본 크기가 작아 정규성을 가정하기 어려울 때 T-검정의 대안으로 사용됩니다.

`scipy.stats.wilcoxon(x, y=None, alternative='two-sided')`

- **x:** 첫 번째 표본 데이터 (또는 단일 표본 데이터)
- **y:** 두 번째 표본 데이터 (대응표본 검정 시). 지정하지 않으면 단일표본 검정을 수행합니다.

- **alternative**: 대립가설의 종류
  - 'two-sided'(양측 검정, 기본값)
  - 'greater'(단측 검정)
  - 'less'(단측 검정).

## 1. 단일표본 Wilcoxon 부호-순위 검정

- **문제**: 어떤 쿠키 제품의 무게 중앙값은 50g으로 알려져 있다. 새로 생산된 쿠키 10개의 무게를 측정했더니 [48, 51, 52, 47, 55, 46, 53, 49, 54, 45] 이었다. 이 쿠키들의 무게 중앙값이 50g과 다르다고 할 수 있는가?

```
from scipy.stats import wilcoxon, shapiro

sample_weights = [48, 51, 52, 47, 55, 46, 53, 49, 54, 45]
pop_median = 50

# 정규성 검정 (표본이 작아 정규성을 가정하기 어려움)
print(f"Shapiro p-value: {shapiro(sample_weights).pvalue:.4f}") # 0.6769

# 단일표본 Wilcoxon 검정 수행
# 검정은 (x - pop_median)에 대해 수행됨
statistic, p_value = wilcoxon([w - pop_median for w in sample_weights])
# 또는 scipy 1.7.0+ 에서는 다음과 같이 직접 수행 가능
# statistic, p_value = wilcoxon(sample_weights, y=None, alternative='two-sided') #
# 단, 이 경우 중앙값이 0인지 검정

print("\n--- One-sample Wilcoxon Signed-rank Test ---")
print(f"Statistic: {statistic:.4f}")
print(f"P-value: {p_value:.4f}")

if p_value < 0.05:
    print("귀무가설 기각: 쿠키 무게의 중앙값은 50g과 유의미하게 다릅니다.")
else:
    print("귀무가설 기각 실패: 쿠키 무게의 중앙값은 50g과 다르다고 할 수 없습니다.")
...

--- One-sample Wilcoxon Signed-rank Test ---
Statistic: 27.5000
P-value: 1.0000
귀무가설 기각 실패: 쿠키 무게의 중앙값은 50g과 다르다고 할 수 없습니다.
...
```

- **결과 해석**
  - **p-value(1.0000)**가 0.05보다 크므로, 귀무가설을 기각하지 못합니다.
  - 즉, 이 표본만으로는 쿠키 무게의 중앙값이 50g과 다르다고 말할 충분한 근거가 없습니다.

## 2. 대응표본 Wilcoxon 부호-순위 검정

- **문제**: 새로운 스트레스 완화 프로그램의 효과를 알아보기 위해, 10명의 참가자를 대상으로 프로그램 참가 전과 후의 스트레스 지수를 측정했다. 프로그램이 스트레스 지수를 낮추는 데 유의미한 효과가 있었는가?

```

stress_before = [8, 7, 9, 6, 8, 7, 9, 5, 8, 7]
stress_after = [6, 5, 7, 5, 6, 6, 8, 4, 7, 5]

# 차이값의 정규성 검정
differences = [b - a for b, a in zip(stress_before, stress_after)]
print(f"Shapiro p-value on differences: {shapiro(differences).pvalue:.4f}") #
0.0003, 정규성 불만족 가정

# 대응표본 Wilcoxon 검정 수행
# 대립가설: before > after (차이가 양수), 즉 프로그램이 효과가 있다.
# wilcoxon(x, y)는 (x-y)의 중앙값이 0인지 검정.
# H1: before > after => before - after > 0 이므로 alternative='greater'
statistic, p_value = wilcoxon(stress_before, stress_after, alternative='greater')

print("\n--- Paired-samples Wilcoxon Signed-rank Test ---")
print(f"Statistic: {statistic:.4f}")
print(f"P-value: {p_value:.4f}")

if p_value < 0.05:
    print("귀무가설 기각: 프로그램은 스트레스 지수를 유의미하게 낮췄습니다.")
else:
    print("귀무가설 기각 실패: 프로그램의 효과가 유의미하지 않습니다.")
...

--- Paired-samples Wilcoxon Signed-rank Test ---
Statistic: 55.0000
P-value: 0.0010
귀무가설 기각: 프로그램은 스트레스 지수를 유의미하게 낮췄습니다.
...

```

#### • 결과 해석

- $p\text{-value}(0.0010)$ 가 0.05보다 작으므로, 귀무가설을 기각합니다.
- 즉, 스트레스 완화 프로그램은 참가자들의 스트레스 지수를 통계적으로 유의미하게 낮추는 효과가 있다고 결론 내릴 수 있습니다.

## 장단점 및 대안

장점	단점
<b>분포에 대한 가정이 없음</b> 데이터가 정규분포를 따르지 않아도 사용할 수 있어 적용 범위가 넓습니다.	<b>검정력(Power) 저하</b> 데이터가 실제로 정규분포를 따를 경우, 대응표본 T-검정에 비해 검정력이 낮습니다. 즉, 실제 차이가 있어도 이를 발견하지 못할 가능성이 T-검정보다 높습니다.
<b>이상치에 강건함(Robust)</b> 실제 값 대신 순위를 사용하므로, 극단적인 이상치의 영향을 덜 받습니다.	<b>정보 손실</b> 실제 데이터 값의 크기 정보를 순위로 변환하는 과정에서 일부 정보가 손실됩니다.

**장점****단점****순서형 데이터에 적용 가능**

데이터가 순위 형태로 주어졌을 때 사용할 수 있습니다.

**대안:****• 부호 검정 (Sign Test)**

- Wilcoxon 부호-순위 검정보다 더 간단한 비모수 검정입니다.
- 차이값의 크기(순위)는 무시하고 오직 부호(+ 또는 -)만을 사용하여 검정합니다.
- 따라서 검정력이 Wilcoxon 검정보다 더 낮지만, 데이터에 대한 가정이 거의 없어 매우 강건합니다.