

맨-휘트니 U 검정 (Mann-Whitney U Test)

- ****독립표본 T-검정(Independent t-test)**의 비모수(non-parametric) 버전**
- 서로 독립적인 두 집단(그룹)의 **중앙값(median)** 또는 **분포**가 동일한지를 검정
- **데이터의 분포에 대한 가정이 필요 없음**
 - 모수 검정인 T-검정은 각 집단의 데이터가 정규분포를 따른다는 가정이 필요

검정 원리

1. 두 집단의 모든 데이터를 하나의 그룹으로 합칩니다.
2. 합쳐진 전체 데이터에 대해 크기 순서대로 순위(rank)를 매깁니다. (가장 작은 값부터 1, 2, 3, ...)
3. 각 집단별로 순위의 합(Rank Sum)을 계산합니다.
4. 만약 두 집단의 분포가 동일하다면, 두 집단의 순위 합은 비슷할 것입니다. 반면, 한 집단이 다른 집단보다 전반적으로 큰 값을 갖는다면, 그 집단의 순위 합이 더 클 것입니다.
5. 맨-휘트니 U 검정은 이 순위 합을 이용하여 검정통계량 U를 계산하고, 이를 기반으로 p-value를 구하여 가설을 검정합니다.

가설 설정

- **귀무가설 (H_0):** 두 집단의 분포는 동일하다. (따라서 중앙값도 같다)
- **대립가설 (H_1):** 두 집단의 분포는 다르다.
 - 한 집단의 값이 다른 집단보다 전반적으로 크거나 작다

적용 가능한 상황

- **독립표본 T-검정의 정규성 가정이 만족되지 않을 때:** 두 개의 다른 반 학생들의 시험 성적, 두 가지 다른 약물을 투여받은 환자 그룹의 반응 시간 등을 비교할 때, 각 그룹의 데이터가 정규분포를 따르지 않는 경우에 사용합니다.
- **데이터가 순서형(Ordinal) 척도일 때:** 데이터가 실제 값은 아니지만 순위 정보(e.g., 만족도 점수)를 가질 때, 두 그룹 간의 차이를 비교하는 데 사용할 수 있습니다.
- **표본 크기가 매우 작을 때:** 표본 크기가 작아 정규성을 가정하기 어려울 때 T-검정의 대안으로 사용됩니다.
- **이상치에 민감하지 않은 검정이 필요할 때:** 실제 값 대신 순위를 사용하므로, 극단적인 이상치의 영향을 덜 받습니다.

`scipy.stats.mannwhitneyu(x, y, alternative='two-sided')`

- **x:** 첫 번째 표본 데이터
- **y:** 두 번째 표본 데이터
- **alternative:** 대립가설의 종류
 - **'two-sided'** (양측 검정, 기본값)
 - **'greater'** (x의 분포가 y보다 큼)
 - **'less'** (x의 분포가 y보다 작음)

예시

- **문제:** 두 가지 다른 교수법(A, B)으로 학생들을 가르친 후, 시험 점수를 비교했다. 두 교수법 간에 학생들의 점수 중앙값에 유의미한 차이가 있는가? (데이터가 정규분포를 따르지 않는다고 가정)

```
from scipy.stats import mannwhitneyu, shapiro

# 표본 데이터 생성
# A 교수법은 점수가 약간 높고, B 교수법은 분산이 더 큰 비정규 데이터
np.random.seed(1)
teaching_method_A = [85, 88, 79, 92, 84, 95, 80, 91, 87, 83]
teaching_method_B = np.random.chisquare(df=10, size=12) * 5 + 30 # 비정규 데이터

# 정규성 검정
print(f"Method A Shapiro p-value: {shapiro(teaching_method_A).pvalue:.4f}") # 0.9091
print(f"Method B Shapiro p-value: {shapiro(teaching_method_B).pvalue:.4f}") # 0.0241, 정규성 불만족

# 맨-휘트니 U 검정 수행
# H1: 두 그룹의 분포가 다르다 (양측 검정)
statistic, p_value = mannwhitneyu(teaching_method_A, teaching_method_B,
                                   alternative='two-sided')

print("\n--- Mann-Whitney U Test ---")
print(f"U-statistic: {statistic:.4f}")
print(f"P-value: {p_value:.4f}")

if p_value < 0.05:
    print("귀무가설 기각: 두 교수법에 따른 점수 분포는 유의미하게 다릅니다.")
else:
    print("귀무가설 기각 실패: 두 교수법에 따른 점수 분포는 차이가 없습니다.")
...

--- Mann-Whitney U Test ---
U-statistic: 80.0000
P-value: 0.1985
귀무가설 기각 실패: 두 교수법에 따른 점수 분포는 차이가 없습니다.
...

# 단측 검정 예시
# H1: A 교수법의 점수가 B 교수법보다 높다 (A > B)
stat_greater, p_greater = mannwhitneyu(teaching_method_A, teaching_method_B,
                                         alternative='greater')
print("\n--- One-sided Test (A > B) ---")
print(f"P-value: {p_greater:.4f}")
...

--- One-sided Test (A > B) ---
P-value: 0.0993
...
```

- **결과 해석:**
 - 정규성 검정

- B 교수법 데이터의 p-value가 0.05보다 작아 정규성 가정을 만족하지 않으므로, 독립표본 T-검정 대신 맨-휘트니 U 검정을 사용하는 것이 적절합니다.
- 양측 검정
 - `p-value(0.1985)`가 0.05보다 크므로, 귀무가설을 채택합니다.
 - 즉, 두 교수법에 따른 학생들의 점수 분포(중앙값)에는 통계적으로 유의미한 차이가 없다고 결론 내릴 수 있습니다.
- 단측 검정
 - `alternative='greater'`로 설정한 검정의 `p-value(0.0993)`가 0.05보다 크므로, A 교수법의 점수가 B 교수법의 점수보다 통계적으로 유의미한 차이가 없다고 결론 내릴 수 있습니다.

장단점 및 대안

장점	단점
분포에 대한 가정이 없음 데이터가 정규분포를 따르지 않아도 사용할 수 있어 적용 범위가 넓습니다.	검정력(Power) 저하 데이터가 실제로 정규분포와 등분산성을 만족할 경우, 독립표본 T-검정에 비해 검정력이 낮습니다. 즉, 실제 차이가 있어도 이를 발견하지 못할 가능성이 T-검정보다 높습니다.
이상치에 강건함(Robust) 실제 값 대신 순위를 사용하므로, 극단적인 이상치의 영향을 덜 받습니다.	정보 손실 실제 데이터 값의 크기 정보를 순위로 변환하는 과정에서 일부 정보가 손실됩니다.
순서형 데이터에 적용 가능 데이터가 순위 형태로 주어졌을 때도 사용할 수 있습니다.	분포의 형태가 다를 때 두 집단의 분산이나 분포의 모양이 심하게 다를 경우, 이 검정은 중앙값의 차이가 아닌 분포 자체의 차이를 검정하는 경향이 있어 해석에 주의가 필요합니다.

대안:

- **Kolmogorov-Smirnov (K-S) 2-표본 검정**
 - 두 표본이 동일한 분포에서 추출되었는지를 검정하는 또 다른 비모수 방법
 - 맨-휘트니 U 검정이 주로 분포의 위치(중앙값) 차이에 민감한 반면, K-S 검정은 분포의 위치, 형태, 퍼짐 등 모든 종류의 차이에 민감합니다.