

# Z-검정 (Z-test)

**Z-검정(Z-test):** 모집단의 분산( $\sigma^2$ )이나 표준편차( $\sigma$ )를 **알고 있을 때**, 표본 데이터(sample)를 이용하여 모집단의 평균( $\mu$ )을 검정하는 통계적 방법

- T-검정과 매우 유사하지만, 모집단의 분산을 아느냐 모르느냐가 두 검정을 구분하는 핵심적인 차이점
- Z-검정은 검정통계량으로 Z-분포(표준정규분포)를 사용

## Z-검정 vs T-검정

구분	Z-검정 (Z-test)	T-검정 (t-test)
핵심 조건	모집단의 분산( $\sigma^2$ )을 <b>알고 있다</b> .	모집단의 분산( $\sigma^2$ )을 <b>모른다</b> .
표본 크기	일반적으로 표본 크기(n)가 클 때 ( $n \geq 30$ ) 사용.	표본 크기(n)가 작을 때 ( $n < 30$ ) 주로 사용.
검정 분포	Z-분포 (표준정규분포)	t-분포

## 실제 적용가능한 상황

실제 데이터 분석 상황에서는 모집단의 분산( $\sigma^2$ )을 정확히 아는 경우가 거의 없습니다. 따라서 Z-검정은 이론적으로는 중요하지만 실제 데이터 분석에서는 T-검정보다 훨씬 드물게 사용됩니다.

다만, **표본의 크기(n)가 충분히 크다면(보통  $n \geq 30$ )**, 중심 극한 정리(CLT)에 의해 표본의 분산( $s^2$ )이 모집단의 분산( $\sigma^2$ )에 매우 가깝게 근사하고, t-분포 또한 표준정규분포에 근사하게 됩니다. 이러한 이유로, 표본 크기가 매우 클 때는 T-검정과 Z-검정의 결과가 거의 동일해지며, 이런 상황에서 Z-검정을 사용하기도 합니다.

특히, 두 집단의 **비율**을 비교하는 검정에서는 Z-검정이 널리 사용됩니다.

## Z-검정의 종류

1. **단일표본 Z-검정 (One-sample Z-test):** 하나의 표본 그룹의 평균이 특정 기준값( $\mu_0$ )과 같은지를 검정합니다.
2. **독립표본 Z-검정 (Two-sample Z-test):** 서로 독립적인 두 표본 그룹의 평균이 같은지를 검정합니다.

## 1. 단일표본 Z-검정 (One-sample Z-test)

- **문제:** 어떤 고등학교 학생들의 평균 IQ는 100, 표준편차는 15로 알려져 있다(모집단 정보). 이 학교에서 특별 교육 프로그램을 받은 30명의 학생들의 평균 IQ를 측정했더니 105였다. 이 특별 교육 프로그램이 학생들의 IQ에 유의미한 영향을 미쳤는가? (유의수준  $\alpha=0.05$ )
- **함수:** `statsmodels.stats.weightstats.ztest(x1, value)`

```
from statsmodels.stats.weightstats import ztest
import numpy as np

# 표본 데이터 생성 (실제로는 측정된 데이터 사용)
np.random.seed(0)
sample_iq = np.random.normal(loc=105, scale=15, size=30)
```

```

# 모집단 평균 (기준값)
pop_mean = 100

# 단일표본 z-검정 수행
# value: 귀무가설에서의 평균 ( $\mu_0$ )
# ddof=0: 모집단 표준편차를 사용한다는 의미 (기본값)
z_statistic, p_value = ztest(sample_iq, value=pop_mean)

print("--- One-sample Z-test ---")
print(f"Z-statistic: {z_statistic:.4f}")
print(f"P-value: {p_value:.4f}")

if p_value < 0.05:
    print("귀무가설 기각: 프로그램은 IQ에 유의미한 영향을 미쳤습니다.")
else:
    print("귀무가설 기각 실패: 프로그램의 영향이 유의미하지 않습니다.")
...

--- One-sample Z-test ---
Z-statistic: 3.8637
P-value: 0.0001
귀무가설 기각: 프로그램은 IQ에 유의미한 영향을 미쳤습니다.
...

```

### • 결과 해석

- p-value(0.0001)가 0.05보다 작으므로, 귀무가설을 기각합니다.
- 즉, 특별 교육 프로그램은 학생들의 IQ에 통계적으로 유의미한 영향을 미쳤다고 결론 내릴 수 있습니다.

## 2. 독립표본 Z-검정 (Two-sample Z-test)

- **문제:** A 도시와 B 도시의 성인 남성 평균 키를 비교하고자 한다. A 도시 남성의 평균 키는 175cm( $\sigma=5$ ), B 도시 남성의 평균 키는 173cm( $\sigma=6$ )로 알려져 있다(모집단 정보). A 도시에서 50명, B 도시에서 60명을 표본으로 추출했을 때, 두 도시의 평균 키에 유의미한 차이가 있는가?
- **함수:** `statsmodels.stats.weightstats.ztest(x1, x2)`

```

# 표본 데이터 생성
np.random.seed(1)
sample_a_height = np.random.normal(loc=175, scale=5, size=50)
sample_b_height = np.random.normal(loc=173, scale=6, size=60)

# 독립표본 z-검정 수행
# value=0: 두 평균의 차이가 0이라는 귀무가설
z_statistic, p_value = ztest(sample_a_height, sample_b_height, value=0)

print("--- Two-sample Z-test ---")
print(f"Z-statistic: {z_statistic:.4f}")
print(f"P-value: {p_value:.4f}")

if p_value < 0.05:

```

```

    print("귀무가설 기각: 두 도시의 평균 키는 유의미하게 다릅니다.")
else:
    print("귀무가설 기각 실패: 두 도시의 평균 키는 차이가 없습니다.")
'''

--- Two-sample Z-test ---
Z-statistic: 1.2027
P-value: 0.2291
귀무가설 기각 실패: 두 도시의 평균 키는 차이가 없습니다.
'''

```

#### • 결과 해석

- p-value(0.2291)가 0.05보다 작으므로, 귀무가설을 채택합니다.
- 즉, A 도시와 B 도시 성인 남성의 평균 키는 통계적으로 유의미한 차이가 없다고 결론 내릴 수 있습니다.

### 3. 두 집단 비율 Z-검정 (Z-test for two proportions)

- **문제:** 새로운 광고 캠페인 전후의 제품 인지도를 비교하고자 한다. 캠페인 전에는 1000명 중 150명(15%)이 제품을 인지했고, 캠페인 후에는 1200명 중 240명(20%)이 제품을 인지했다. 광고 캠페인이 제품 인지도에 유의미한 향상을 가져왔는가?
- **함수:** `statsmodels.stats.proportion.proportions_ztest(count, nobs)`

```

from statsmodels.stats.proportion import proportions_ztest

# 데이터 설정
count = np.array([150, 240]) # 각 그룹에서 성공(인지)한 횟수
nobs = np.array([1000, 1200]) # 각 그룹의 전체 관측(표본) 수

# 두 집단 비율 Z-검정 수행
z_statistic, p_value = proportions_ztest(count, nobs, value=0)

print("--- Z-test for two proportions ---")
print(f"Z-statistic: {z_statistic:.4f}")
print(f"P-value: {p_value:.4f}")

if p_value < 0.05:
    print("귀무가설 기각: 광고 캠페인 전후의 인지도는 유의미하게 다릅니다.")
else:
    print("귀무가설 기각 실패: 인지도 차이가 유의미하지 않습니다.")
'''

--- Z-test for two proportions ---
Z-statistic: -3.0577
P-value: 0.0022
귀무가설 기각: 광고 캠페인 전후의 인지도는 유의미하게 다릅니다.
'''

```

#### • 결과 해석

- p-value(0.0022)가 0.05보다 작으므로, 귀무가설을 기각합니다.
- 즉, 광고 캠페인은 제품 인지도를 통계적으로 유의미하게 향상시켰다고 결론 내릴 수 있습니다.

