

# R을 이용한 통계 기초와 데이터 분석

남현진

한성대학교

2020

## T test

두 집단의 평균의 차이 유무를 비교하는 테스트이다. Z test의 경우 모집단의 표준편차를 알고 있어야 하지만 현실에서는 모집단에 대한 정확한 평균이나 분산에 대한 통계량을 알지 못하는 경우가 대부분이다. 따라서 Z 검정 보다는 조금 더 일반적인 경우에도 이용할 수 있는 T 검정을 많이 사용한다.

## T test 가정

Student T 검정을 위해서는 두가지 가정이 만족되어야한다. T test에서도 여러가지 종류가 있는데 만약 등분산성 가정이 만족되면 Student T test를 사용하고 등분산성이 만족되지 않는다면 Welch T-Test를 이용해야 한다.

### Assumptions

- 정규성 가정 (Normality Assumption)
- 등분산성 가정 (Homogeneity of variance) (2 표본일 경우)

## 정규성 가정

일반적으로 표본의 수가 30개 이상이면 중심극한 정리에 의해 검정 없이도 정규성을 가정할 수 있다. 하지만 표본의 개수가 30개 이하라면 Shapiro-Wilk test, Pearson's chi-squared test 등의 방법을 이용하여 표본이 정규분포를 따름을 증명해야한다. 알파를 0.05라고 두면 p-value가 0.05보다 크면 자료가 정규성을 만족한다고 말할 수 있다.

### Assumptions

- 귀무가설  $H_0$ : 자료는 정규분포를 따른다.
- 대립가설  $H_1$ : 자료는 정규분포를 따르지 않는다.

## 등분산성 가정

등분산성은 분석하는 두 표본의 분산이 같다는 뜻이다. 만약 두 표본의 분산이 같다면 Student T test를 이용하고 두 표본의 분산이 같지 않다면 Welch T test를 이용해야 한다. R 에서 Student T test는 `t.test` 함수의 `var.equal = F`를 통해 사용할 수 있고 Student T Test는 `var.equal = T`를 통해 사용할 수 있다. 알파를 0.05라고 두면 p-value가 0.05보다 크면 자료가 등분산성을 만족한다고 말할 수 있다.

### Assumptions

- 귀무가설  $H_0$ : 두 자료의 분산은 같다.
- 대립가설  $H_1$ : 두 자료의 분산은 같지 않다.

## 일표본 T 검정

어떤 전구의 지속 시간이 39000분으로 알려져 있다고 한다. 이 때 75개의 전구 표본을 뽑아 지속시간을 알아보니 평균 36500분과 표준편차 2000분으로 나타났다고 한다. 이 때 이 전구는 평균 지속시간이 39000분이라 말할 수 있을까?

- 귀무가설  $H_0$ : 전구의 지속시간은 39000분이다.
- 대립가설  $H_1$ : 전구의 지속시간은 39000분이 아니다.

## 일표본 T 검정

### Result

- t-test의 검정통계량  $t = -10.565$ 이다.
- 자유도  $df$ 는 74이다.
- 유의확률  $p\text{-value}$ 는  $< 2.2e-16$  이다.
- 전구의 지속 시간의 95 % 신뢰구간은 [36318.76 37169.64]이다.
- 표본의 평균은 36744.2이다.

$p$  value는  $< 2.2e-16$ 로 0.05보다 작다. 따라서 전구의 지속시간은 39000분이라는 귀무가설을 기각한다.

## 독립표본 T 검정

9명의 여성 집단과 9명의 남성 집단이 있다. 모든 사람들의 몸무게를 측정했을 때 남성 집단의 몸무게의 평균과 여성 집단의 몸무게 평균을 비교해보려 한다. 이 때 남성 집단의 몸무게가 여성 집단의 몸무게와 같다고 말할 수 있을까?

- 귀무가설  $H_0$ : 남성과 여성의 몸무게는 같다.
- 대립가설  $H_1$ : 남성 집단과 여성 집단의 몸무게는 다르다.



## 독립표본 T 검정

### Result

- t-test의 검정통계량  $t = -2.784$ 이다.
- 자유도  $df$ 는 16이다.
- 유의확률 p-value는 0.01327이다.
- 두 집단의 평균 차이의 95% 의 신뢰구간은  $[-29.748019, -4.029759]$ 이다.
- 두 집단의 평균은 다음과 같다.  $\bar{x} = 52.1$  ,  $\bar{y} = 68.99$

p value는 0.01327로 0.05보다 작다. 따라서 남성의 평균 몸무게는 여성의 평균 몸무게와 같다는 귀무가설을 기각하고 두 집단의 몸무게가 다르다라고 결론내릴 수 있다.

## 대응표본 T 검정

어떤 학교의 1000명의 학생들의 중간고사 수학 성적의 평균은 64점 표준편차는 10이라고 알려져 있다. 이 때 이 학교에 새로운 교육 방식이 적용되었고 기말고사를 치르게 되었다. 이 때 기말고사 수학 성적의 평균은 68점 표준편차는 20점이라고 했을 때 새로운 교육 방식 이후의 학생들의 점수가 통계적으로 유의미하게 올랐다고 말할 수 있을까?

- 귀무가설  $H_0$ : 새로운 교육 방식의 전과 후의 학생들의 성적은 같다.
- 대립가설  $H_1$ : 새로운 교육 방식의 적용 후 학생들의 성적은 상향되었다.

## 대응표본 T 검정

### Result

- t-test의 검정통계량  $t = -3.8091$ 이다.
- 자유도  $df$ 는 9990이다.
- 유의확률  $p$ -value는  $7.4e-05$ 이다.
- 시험 전과 후의 점수 차이의 95%의 신뢰구간은  $[-Inf, -1.5848]$ 이다.
- 두 집단의 평균의 차이는  $-2.791243$ 이다.

$p$  value는  $7.4e-05$ 로 0.05보다 작다. 따라서 새로운 교육 방식의 전과 후의 학생들의 성적은 같다는 귀무가설을 기각하고 새로운 교육 방식의 적용 후 학생들의 성적은 상향되었다고 결론내릴 수 있다.