

# R을 이용한 통계 기초(4일차)

인하대학교 대학원 통계학과 국성희

# Contents

1. 통계적 가설검정
2. T-검정
3. 카이제곱 검정
4. 연습문제



# Contents



# 1. 통계적 가설검정

- 1.1 통계적 가설검정

대립가설( $H_1$ ) : 입증하여 주장하고자 하는 가설

귀무가설( $H_0$ ) : 대립가설의 반대 가설로 대립가설을 입증할 수 없을 때 대립가설을 무효화 시키면서 받아들이는 가설

P-값(유의확률) : 주어진 검정통계량의 관측치로부터  $H_0$ 를 기각하게 하는 최소의 유의 수준

검정 통계량 :  $H_0$ 를 기각하거나  $H_0$ 를 기각하지 않고 유지하는 결정을 내리는데 사용되는 표본의 함수

# 1. 통계적 가설검정

- 1.1 통계적 가설검정

통계적 결정	귀무가설 $H_0$	
	참	거짓
$H_0$ 를 기각함	제 1종 오류 $\alpha$	옳은 결정 $1-\beta$
$H_0$ 를 채택함	옳은 결정 $1-\alpha$	제 2종 오류 $\beta$

$\alpha = P[\text{제 1종 오류}] = P[H_0 \text{가 참일 때 } H_0 \text{를 기각}]$

$\beta = P[\text{제 2종 오류}] = P[H_0 \text{가 거짓일 때 } H_0 \text{를 채택}]$

$1-\beta = P[H_0 \text{가 거짓일 때 } H_0 \text{를 기각}] = \text{검정력(power)}$

## 2. T-검정

- 2.1 일집단 t-검정

다음은 암컷 원숭이의 몸무게를 조사한 자료이다(단위 kg). 원숭이 몸무게는 정규분포를 따른다고 할 수 있다고 한다. 암컷 원숭이의 평균 몸무게가 8.5kg 이라고 할 수 있는지 검정하고자한다.

8.30 9.50 9.60 8.75 8.40 9.10 8.15 8.80

```
x=c(8.30, 9.50, 9.60, 8.75, 8.40, 9.10, 8.15, 8.80)
```

```
t.test(x,mu=8.5)
```

## 2. T-검정

- 2.2 이집단 t-검정

- $\sigma_1^2 = \sigma_2^2 = \sigma^2$ 이며  $\sigma^2$ 를 모르는 경우

다음은 지혈제 A와 B의 효과를 측정한 데이터이다. 실험 참여자에게 같은 상처를 낸 후 지혈될 때까지의 시간을 측정하였으며 모집단은 정규분포를 따른다고 한다.

지혈제 A와 B의 효과가 차이가 있는지 검정하고자 한다.

	지혈시간				
지혈제A	1.1	2.3	4.3	2.2	5.3
지혈제B	2.3	4.3	3.5		

`x1=c(1.1,2.3,4.3,2.2,5.3)`

`x2=c(2.3,4.3,3.5)`

`t.test(x1,x2,var.equal=TRUE, alternative="two.sided")`

## 2. T-검정

- 2.2 이집단 t-검정

#두 집단의 분산이 다르다고 가정 할 경우 (양측검정)

```
t.test(x1,x2,var.equal=FALSE, conf.level=0.95)
```

#두 집단의 분산이 같다고 가정 할 경우 (단측1)

```
t.test(x1,x2,var.equal=TRUE, alternative="greater")
```

#두 집단의 분산이 같다고 가정 할 경우(단측2)

```
t.test(x1,x2,var.equal=TRUE, alternative= " less")
```



## 2. T-검정

- 2.2 이집단 t-검정

다음은 두 가지 배양법에 따라 호박잎의 질소 성분 함량을 측정한 데이터이며, 정규분포를 따른다고 가정할 수 있다고 한다. 두 배양법 간에 질소성분함량평균 차이가 있는지 검정하고자 한다.

(method.txt file)

```
met=read.table("file 위치",header=T)
```

```
#두 집단의 분산이 같다고 할 경우
```

```
t.test(x~method,var.equal=T, data=met)
```

```
#두 집단의 분산이 다르다고 한 경우
```

```
t.test(x~method,var.equal=F, data=met)
```

## 2. T-검정

- 2.3 이집단 분산비 F-검정

지혈제A와 지혈제B의 분산이 동일한가에 대해 검정하고자 한다.

`var.test(x1,x2)`

## 2. T-검정

- 2.4 짝지어진 표본에 대한 t-검정

각 개체에 대해 정규분포를 따르는 모집단으로부터 쌍으로 측정하여 확률표본을 얻은 경우 두 집단의 모평균이 같은지 검정하고자 하는 경우

A대학에서 10명의 학생을 대상으로 학습법 강좌 수강 전후의 B과목 시험 점수가 정리되어있다. 강좌 수강 전후 점수 차이가 있는가를 검정하여라.

시험	점수									
pre test	77	56	64	60	58	72	67	78	67	79
post test	99	80	78	65	59	67	65	85	74	80

## 2. T-검정

- 2.4 짝지어진 표본에 대한 t-검정

```
pre=c(77, 56, 64, 60, 58, 72, 67, 78, 67, 79)
```

```
post=c(99, 80, 78, 65, 59, 67, 65, 85, 74, 80)
```

```
t.test(post,pre,paired=T)
```

```
Paired t-test
```

```
data: post and pre
```

```
t = 2.3906, df = 9, p-value = 0.04052
```

```
alternative hypothesis: true difference in means is not equal to 0
```

```
95 percent confidence interval:
```

```
0.3974552 14.4025448
```

```
sample estimates:
```

```
mean of the differences
```

```
7.4
```

### 3. 카이제곱 검정

- 3.1 카이제곱 독립성 검정

2개의 범주형 변수간에 관련성에 대해 알아보기 위해  
두 변수간의 독립성여부에 대한 검정을 하고자 할 때  
카이제곱 검정(chi-square test)을 한다.

```
x=matrix(c(54,63,45,65),nrow=2)
```

```
chi=chisq.test(x)
```

```
#각 칸의 기대도수와 잔차를 보고자 하는 경우
```

```
chi$expected
```

```
chi$residuals
```

	고등학교 졸업여부	
	Yes	No
고수입	54	45
저수입	63	65

### 3. 카이제곱 검정

- 3.1 카이제곱 독립성 검정

데이터가 개체에 대한 정보로 주어진 경우

```
x=seq(1,12)
```

```
gender=c("M","M","M","M","M","M","F","F","F","F","F","F")
```

```
admit=c("Y","Y","Y","N","N","Y","Y","Y","N","N","Y","Y")
```

```
y=cbind(x,gender,admit)
```

```
table(gender,admit)
```

```
chisq.test(gender,admit)
```

	x	gender	admit
1	1	M	Y
2	2	M	Y
3	3	M	Y
4	4	M	N
5	5	M	N
6	6	M	Y
7	7	F	Y
8	8	F	Y
9	9	F	N
10	10	F	N
11	11	F	Y
12	12	F	Y

### 3. 카이제곱 검정

- 3.2 카이제곱 적합도 검정

알사탕 100개가 들어있는 한 봉지에서 빨강(1), 노랑(2), 파랑(3), 초록색(4) 각각의 사탕 개수를 세어보니 30, 20, 27, 23 이었다. 4가지 색깔의 분포가 균일분포를 따른다고 할 수 있는가?

```
y=c(30,20,27,23)
```

```
p=c(0.25,0.25,0.25,0.25)
```

```
chisq.test(y,p=p)
```

## 4. 연습문제

- 1. 지난 1년동안 A학과의 각 학년별 학과행사 참석여부를 조사한 데이터 이다. 학년과 참석여부가 서로 독립인지 유의수준 5%에서 검정하시오

	1	2	3	4
참석	40	30	35	20
불참	20	30	45	40

- 2. 2003년 미국 미시간 주에서 일어난 살인 사건수를 요일별로 정리한 결과이다. 요일별 살인 사건 발생 확률이 같다고 할 수 있는지 유의 수준 5%에서 검정하시오

일요일	월요일	화요일	수요일	목요일	금요일	토요일
53	42	51	45	36	37	65



## 4. 연습문제

- 3. 전기기구를 고치는데 까지 걸리는 시간은 정규분포를 따른다고 할 수 있으며 아래와 같은 10개의 가구 수리시간에 대한 데이터를 얻었다. (단위 : 시간)

159 280 101 121 224 222 379 179 250 170

- (a) 평균 수리 시간이 225시간을 넘는다고 할 수 있는가
- (b) 평균 수리시간에 대한 95% 신뢰구간을 구하십시오
- (c) 평균 수리 시간에 대한 98%신뢰 구간을 구하십시오

- 4. 어려운 것을 기억하는 두 가지 방법이 있는데 어떤 방법이 더 효과적인가를 결정하고자 한다. IQ와 성적을 기준으로 해서 짝지은 9쌍의 학생들을 대상으로 각 쌍에서 임의로 배정하여 기억력 실험을 시행한 결과는 다음과 같다. 효과의 차이는 유의 한가?

A : 90 86 72 65 44 52 46 38 43

B : 85 87 70 62 44 53 42 35 46

## 4. 연습문제

- 4. 어려운 것을 기억하는 두 가지 방법이 있는데 어떤 방법이 더 효과적인가를 결정하고자 한다. IQ와 성적을 기준으로 해서 짝지은 9쌍의 학생들을 대상으로 각 쌍에서 임의로 배정하여 기억력 실험을 시행한 결과는 다음과 같다. 효과의 차이는 유의 한가?

A : 90 86 72 65 44 52 46 38 43

B : 85 87 70 62 44 53 42 35 46

Q&A

