0) 자료의 형태

A-type(2강의)

남자 키	여자 키
179	158
168	165
170	160
180	160
174	160
168	170
	153

B type(3강의)

gender	height
1	179
1	168
1	170
2	158
2	165
1	180
1	174
2	160
1	168
2	160
2	170
2	153

B. 자료의 형태 (B-type)

B-1. 정규모집단, 등분산인 경우 (R 제공)

⑵ 자료2 - 두가지 배양법에 대한 질소성분함량

```
method = c ( 1, 1, 1, 1, 1, 2, 2, 2, 2 )          # 자료입력 2
x = c ( 19.1, 32.8, 27.6, 25.9, 28.5, 17.0, 16.4, 16.8, 15.5)
t.test(x~method, var.equal = T)          # 양적자료~그룹자료 순으로
```

$H_1 : \mu_1 \neq \mu_2$

```
        Two Sample t-test

data:  x by method
t = 4.0647, df = 7, p-value = 0.004781
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  4.331008 16.378992
sample estimates:
mean in group 1 mean in group 2
         26.780          16.425
```

해석 :

- 추정

  : 방법1의 평균 질소성분함량은 26.78, 방법2의 평균 질소성분함량은 16.46이며, 방법1과 방법2의 평균 질소성분함량차에 대한 95% 신뢰구간은 (4.33, 16.4)이다.

- 가설검정

  ① 가설 H_0: 방법1과 방법2의 평균 질소성분함량은 같다.

  H_1: 방법1과 방법2의 평균 질소성분함량은 다르다.

  ② 유의수준 α=0.05

  ③ 검정통계량 T값 = 4.0647

  ④ P값 = 0.0047 < α => H_0를 기각

  ⑤ 결론 : 유의수준 5%에서 방법1과 방법2의 평균 질소성분함량은 다르다고 할 수 있다. 방법1의 평균 질소성분함량(26.78)이 방법2의 평균 질소성분함량(16.37)보다 크다고 할 수 있다.

B-2. 정규모집단, 등분산이 아닌 경우(이분산) (R 제공)

t.test(x~method)

```
        Welch Two Sample t-test

data:  x by method
t = 4.5889, df = 4.1769, p-value = 0.00913
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  4.193073 16.516927
sample estimates:
mean in group 1 mean in group 2 
         26.780          16.425 
```

B-3. 이표본 분산비 F-Test (R 제공)

var.test(x~method) # $H_1 : {\sigma_1}^2 \neq {\sigma_2}^2$

```
        F test to compare two variances

data:  x by method
F = 56.287, num df = 4, denom df = 3, p-value = 0.00748
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
   3.727375 561.699204
sample estimates:
ratio of variances 
          56.28701 
```

해석 :

- 가설검정

① 가설 H_0 : 방법1과 방법2의 분산은 같다.

H_1 : 방법1과 방법2의 분산은 다르다.

② 유의수준 α = 0.05

③ 검정통계량 F값 = 56.287

④ P값 = 0.00748 < α => H_0를 기각

⑤ 결론 : 유의수준 5%에서 방법1과 방법2의 분산은 다르다고 할 수 있다.

=> 이표본 T-Test에서 이분산 가정 적용(B-2)

B-4. 이표본 T-Test (독립표본) (R 프로그래밍)

[NOTE] B-type => A-type 변환

```
method = c ( 1, 1, 1, 1, 1, 2, 2, 2, 2 )
x = c ( 19.1, 32.8, 27.6, 25.9, 28.5, 17.0, 16.4, 16.8, 15.5)

x1 = rep(0, length(method)) ; x1
x2 = rep(0, length(method)) ; x2
for ( i in 1 : length(x)) {
   if ( method[i] < 2) x1[i] = x[i]
   else  x2[i] = x[i]
}
x1
x2

 x1 = x1[ x1 > 0.01]
 x1
 x2 = x2[ x2 > 0.01]
 x2
```

B-type : 이표본 T-Test (독립표본) (R 프로그래밍)

```
T_test_2B = function(data , group) {
x = rep(0, length(group))            # 자료 변환
y = rep(0, length(group))
for ( i in 1 : length(group)) {
   if ( group[i] < 2) x[i] = data[i]
   else  y[i] = data[i]
}
 x = x[ x > 0.01] ; y = y[ y > 0.01]

n1 = length(x) ; n2 = length(y)            # 등분산 검정
s1 = var(x) ; s2 = var(y)
F = s1 / s2
pvalue = min( 2*pf(F, n1-1, n2-1) , 2*( 1 - pf(F, n1-1, n2-1) ) )

cat( "  ============  이표본 분산비 검정  ============", "\n", "\n")
cat( "          F = " , F, " , P - value = " , pvalue , "\n", "\n")

xbar = mean(x) ; ybar = mean(y)                # 등분산인 경우 t-test
sp = sqrt ( ( (n1 - 1) * s1 + (n2 - 1) * s2 ) / ( n1 + n2 - 2) )
T = ( xbar - ybar ) / ( sp * sqrt ( 1/n1 + 1/n2) )
pvalue = 2 *( 1 - pt( abs(T), n1 + n2 - 2) )
cat( "  ============  이표본 평균차 검정  ============", "\n", "\n")
cat( " 등분산인 경우  : T = " , T, " , P - value = " , pvalue , "\n")

df = ( s1/n1 + s2/n2 )^2 / ( (s1/n1)^2 / (n1 - 1) + (s2/n2)^2 / (n2 - 1) )
T = ( xbar - ybar ) / sqrt ( s1/n1 + s2/n2)          # 이분산인 경우 t-test
pvalue = 2 *( 1 - pt( abs(T), df ) )
cat( " 이분산인 경우  : T = " , T, " , P - value = " , pvalue , "\n")
}
T_test_2B(x, method)
```

```
 ============  이표본 분산비 검정  ============

          F =  56.28701  , P - value =  0.007479745

 ============  이표본 평균차 검정  ============

등분산인 경우  : T =  4.064694  , P - value =  0.004781153
이분산인 경우  : T =  4.58886  , P - value =  0.009129666
```

[NOTE] SPSS

method	x
1.00	19.10
1.00	32.80
1.00	27.60
1.00	25.90
1.00	28.50
2.00	17.00
2.00	16.40
2.00	16.80
2.00	15.50

독립표본 검정

		Levene의 등분산 검정		평균의 동일성에 대한 T 검정						
		F	유의확률	t	자유도	유의확률 (양측)	평균차이	표준오차 차이	차이의 95% 신뢰구간 하한	차이의 95% 신뢰구간 상한
x	등분산을 가정함	3.265	.114	4.065	7	.005	10.35500	2.54755	4.33101	16.37899
	등분산을 가정하지 않음			4.589	4.177	.009	10.35500	2.25655	4.19307	16.51693

[NOTE] R에서 Levene Test(참고만)

install.packages("lawstat")

library(lawstat)

levene.test(x, method) **# 디폴트 median base**

```
Modified robust Brown-Forsythe Levene-type test based on the absolute
deviations from the median

data:  x
Test Statistic = 2.3939, p-value = 0.1657
```

levene.test(x, method, location = c("mean")) **# mean base**

```
Classical Levene's test based on the absolute deviations from the mean ( none
not applied because the location is not set to median )

data:  x
Test Statistic = 3.265, p-value = 0.1137
```

(3) 우리자료

var.test(height~gender) **#** $H_1 : \sigma_1^{\ 2} \neq \sigma_2^{\ 2}$ **=> 등분산**

t.test(height~gender, var.equal = T)

T_test_2B(height, gender) **# error ㅜㅜㅜ**

gender **# 남자, 여자 error ㅜㅜㅜ**

gender <- data$gender **# 다시 gender를 불러들임**

T_test_2B(height, gender)

[과제24] (여러분은 따라서 해보시고, 아래 실습문제를 과제로 제출하시길)

- 과제방법 :
  ① R에서 제공 결과 => 결과분석
  ② R 프로그래밍 결과 => ①의 결과와 같음을 확인

[실습3] **우리자료 :** 성별에 따라 몸무게에 차이가 있는지 검정하시오.

[실습4] **우리자료 :** 성별에 따라 BMI에 차이가 있는지 검정하시오.

첨부파일 : 학번이름24.hwp (예 : 20192260홍길동24.hwp)