

1번

다음은 어떤 육상 선수의 100m 달리기 기록을 9번 잰 것이다.

11.1 11.3 10.8 10.6 11.4 11.2 11.0 10.5 11.5

위의 자료로 볼 때, 이 선수의 평균 기록이 11초라고 말할 수 있는가? 유의수준 $\alpha=0.05$ 에서 검정하여라.

```
In [ ]: record = c(11.1, 11.3, 10.8, 10.6, 11.4, 11.2, 11.0, 10.5, 11.5)

t.test(record, mu=11, conf.level=0.95)
```

One Sample t-test

```
data: record
t = 0.38052, df = 8, p-value = 0.7135
alternative hypothesis: true mean is not equal to 11
95 percent confidence interval:
 10.77511 11.31378
sample estimates:
mean of x
 11.04444
```

답변: 이 선수의 평균 기록은 95% 확률로 11초라고 할 수 있다.

95% 신뢰구간은 (10.77511 ~ 11.31378) 이며, 표본 평균은 11.04444로 신뢰구간에 속한다.

2번

서로 다른 두 개의 낚시대를 사용했을 때, 잡히는 고기의 마리수에 차이가 있는지를 알아보기 위하여 다음과 같은 실험을 하였다. 1년에 16일(시기)을 정하여 4년에 걸쳐 같은 장소에서 두 사람이 서로 다른 낚시대를 사용하여 잡은 고기의 마리수를 관측하였다. 아래의 관측값은 각 시기별의 평균값이다. 두 개의 낚시대에 차이가 있다고 말할 수 있는가?

낚시대 A = 6.64 7.89 1.83 0.42 0.85 0.29 0.57 0.63 0.32 0.37 0.00 0.11 4.86 1.80 0.23 0.58

낚시대 B = 6.83 8.21 2.17 0.75 1.61 0.75 0.83 0.56 0.76 0.38 0.48 0.52 5.38 2.33 0.91 0.79

(a) Two-sample test를 이용하여 유의수준 $\alpha=0.01$ 에서 검정하여라.

(b) 공통 분산을 가질 때의 t 검정 방법을 이용하면 어떻게 되는가?

```
In [ ]: # 낚시대 A와 B의 관측값
fish_A <- c(6.64, 7.89, 1.83, 0.42, 0.85, 0.29, 0.57, 0.63, 0.32, 0.37, 0.00, 0.11,
fish_B <- c(6.83, 8.21, 2.17, 0.75, 1.61, 0.75, 0.83, 0.56, 0.76, 0.38, 0.48, 0.52,

# (a) Two-sample test를 이용한 검정 (유의수준=0.01)
t.test(fish_A, fish_B, alternative = "two.sided", conf.level = 0.99)
```

Welch Two Sample t-test

```
data: fish_A and fish_B
t = -0.41969, df = 30, p-value = 0.6777
alternative hypothesis: true difference in means is not equal to 0
99 percent confidence interval:
 -2.770817  2.037067
sample estimates:
mean of x mean of y
 1.711875  2.078750
```

답변: 유의수준 0.01에서 낚시대 A와 B의 평균은 동일하다고 할 수 있다.

99% 신뢰구간은 (-2.770817 ~ 2.037067)이며, 평균의 차는 -0.41969로 신뢰구간에 속한다.

```
In [ ]: # (b) 공통 분산을 가질 때의 t 검정
t.test(fish_A, fish_B, alternative = "two.sided", var.equal = T, conf.level = 0.99)
```

Two Sample t-test

```
data: fish_A and fish_B
t = -0.41969, df = 30, p-value = 0.6777
alternative hypothesis: true difference in means is not equal to 0
99 percent confidence interval:
 -2.770815  2.037065
sample estimates:
mean of x mean of y
 1.711875  2.078750
```

답변: 유의수준 0.01에서 낚시대 A와 B의 평균은 동일하다고 할 수 있다.

99% 신뢰구간은 (-2.770815 ~ 2.037065)이며, 평균의 차는 -0.41969로 신뢰구간에 속한다.

3번

4개 회사 (A1, A2, A3, A4)에서 입하된 전지의 기전력을 비교하기 위하여 각 회사로부터 6개 썬의 기전력(단위는 Volt)을 측정한 결과 다음의 관측값을 얻었다. 이 관측값에 대하여 one-way analysis of variance의 모형을 적용할 때, 회사간에 기전력의 차이가 있는지 유의수준 5%에서 검정하라.

A1 = 1.52, 1.50, 1.52, 1.49, 1.52, 1.53

A2 = 1.55, 1.54, 1.57, 1.56, 1.56, 1.55

A3 = 1.48, 1.47, 1.46, 1.50, 1.48, 1.48

A4 = 1.57, 1.59, 1.60, 1.58, 1.57, 1.59

```
In [ ]: # 데이터 입력
A1 <- c(1.52, 1.50, 1.52, 1.49, 1.52, 1.53)
A2 <- c(1.55, 1.54, 1.57, 1.56, 1.56, 1.55)
A3 <- c(1.48, 1.47, 1.46, 1.50, 1.48, 1.48)
A4 <- c(1.57, 1.59, 1.60, 1.58, 1.57, 1.59)

# 회사별 데이터가 정규분포인지 확인한다.
shapiro.test(A1)
shapiro.test(A2)
```

```
shapiro.test(A3)
shapiro.test(A4)

# 회사별 데이터를 q3 변수에 입력한다.
q3 <- data.frame(voltage = c(A1, A2, A3, A4), corp = rep(c("A1", "A2", "A3", "A4"),

# 회사간에 기전력의 차이가 있는지 검정
anova(lm(q3$voltage ~ q3$corp))
```

Shapiro-Wilk normality test

```
data: A1
W = 0.86626, p-value = 0.2117
Shapiro-Wilk normality test
```

```
data: A2
W = 0.96004, p-value = 0.8201
Shapiro-Wilk normality test
```

```
data: A3
W = 0.92121, p-value = 0.5141
Shapiro-Wilk normality test
```

```
data: A4
W = 0.9067, p-value = 0.415
A anova: 2 × 5
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
	<int>	<dbl>	<dbl>	<dbl>	<dbl>
q3\$corp	3	0.03835	0.01278333	77.47475	3.475136e-11
Residuals	20	0.00330	0.00016500	NA	NA

답변: p-value가 3.475e-11로 매우 작은 값이므로 귀무가설을 기각한다. 즉, 회사간에 기전력의 차이가 있다.

```
In [ ]: pairwise.t.test(q3$voltage, q3$corp, p.adj="bonferroni")
```

Pairwise comparisons using t tests with pooled SD

```
data: q3$voltage and q3$corp
```

```
      A1      A2      A3
A2 0.00010 -      -
A3 0.00079 1.1e-08 -
A4 5.0e-08 0.00642 4.2e-11
```

P value adjustment method: bonferroni

참고로, 모든 회사 조합(A1-A2, A1-A3, A1-A4, A2-A3, A2-A4, A3-A4)에서 귀무가설이 기각됨을 알 수 있다.

4번

어떤 약품의 불순물을 제거하는 정제공정에서 첨가제의 양(A) 5종류, 처리온도(B) 4종류를 택하여 총 20개 조에 실험을 랜덤한 순서로 행한 후 불순물의 잔류량(g)을 측정한 결과 다음과 같았다. two-way analysis of variance의 모형을 적용하여 A, B인자의 영향을 유의수준 5%에서 검정하라.

인자 A/B	B1	B2	B3	B4
A1	26	28	31	36
A2	30	33	27	31
A3	30	32	28	33
A4	29	28	32	30
A5	32	27	31	26

```
In [ ]: gram = c(26, 28, 31, 36,
                 30, 33, 27, 31,
                 30, 32, 28, 33,
                 29, 28, 32, 30,
                 32, 27, 31, 26)

q4 = data.frame(gram = gram, groupA=rep(c("A1", "A2", "A3", "A4", "A5"), each=4), g

aggregate(gram~groupA, q4, mean)
aggregate(gram~groupB, q4, mean)

anova(lm(gram~groupA+groupB, data=q4))
```

A data.frame: 5 ×
2

groupA	gram
<chr>	<dbl>
A1	30.25
A2	30.25
A3	30.75
A4	29.75
A5	29.00

A dataframe: 4 × 2

groupB	gram
<chr>	<dbl>
B1	29.4
B2	29.6
B3	29.8
B4	31.2

A anova: 3 × 5

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
	<int>	<dbl>	<dbl>	<dbl>	<dbl>
groupA	4	7	1.750000	0.1826087	0.9430008
groupB	3	10	3.333333	0.3478261	0.7914310
Residuals	12	115	9.583333	NA	NA

답변) 첨가제의 양 A, 처리온도 B 의 p-value가 모두 0.05보다 큰 값이므로 귀무가설을 기각하지 못한다. 즉, 첨가제의 양과 처리온도는 불순물의 잔류량에 영향을 주지 못한다.