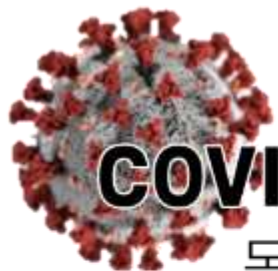


R STUDIO 활용

빅데이터 분석 :

목차 :

1. 독립표본 T 검정
2. 대응표본 T 검정
3. 일원분산분석
4. 반복측정 분석
5. 단순회귀 분석
6. 로지스틱 회귀분석



COVID19

목적 : 코로나가 현대사회에 시사하는 문제점

얻고자 하는 것

: 사회적 거리두기의 중요성, 밀집 지역군의 위험성,

인간이 환경에 미치는 영향, 건강관리의 중요성, 코로나로 인해 변화된 일상생활

R STUDIO 분석 1. Independent sample T-Test (독립표본 T 검정)

문제정의 KOO뉴스는 코로나19로 인해 초미세먼지 농도가 떨어졌다고 발표했다. 이에 따라, 2019년 3월과 2020년 3월의 초미세먼지 농도의 차이가 있다는 연구가설을 세우고, 본 연구가설이 맞는지, 맞으면 얼마나 차이가 나는지 확인해보도록 한다.

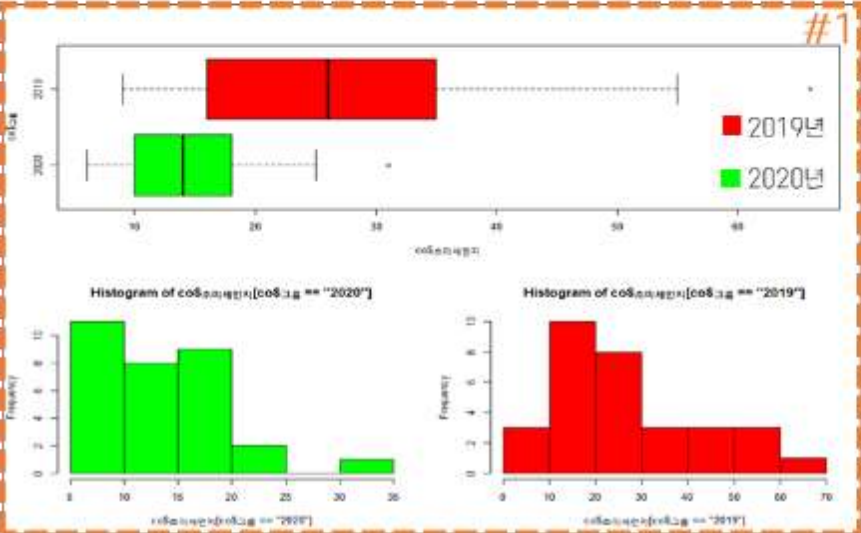
▶ 데이터 불러오기

```
> # 01. 데이터 불러오기
> co <- read.csv("초미세.csv",
+               header=TRUE,
+               na.strings = ".")
```

▶ 기본 통계치 확인

```
> describeBy(co$초미세먼지,
+            co$그룹,
+            mat=T)
+
+   item group1 vars  n   mean    sd median trimmed  mad min max
+ X11    1    2020    1 31 14.41935  6.10341  14 13.92  5.9304  6 31
+ X12    2    2019    1 31 27.90323 15.29129  26 26.44 16.3086  9 66
+
+ range      skew      kurtosis      se
+ X11    25 0.6516209 -0.04109016 1.096205
+ X12    57 0.7990436 -0.42509914 2.746392
```

▶ Boxplot, Histogram을 통한 데이터 분석



해석

- #1 Boxplot과 Histogram을 통해 2019년에 비해 2020년도의 초미세먼지 농도가 감소하였음을 알 수 있다.
- #2 등분산 검정에서 P값이 0.05보다 작기 때문에, 이분산이다.
- #3 T 검정에서 P값이 0.05보다 작기 때문에, 2019년과 2020년도의 초미세먼지 농도는 차이가 있다는 연구가설을 채택하였다.

결론

P값이 0.00004898로, 2019년 3월과 2020년 3월의 초미세먼지 농도 차이는 통계적으로 유의한 차이가 있다는 것을 알 수 있다.

▶ 등분산 검정 (var.test)

```
> var.test(co$초미세먼지 ~ co$그룹, data = co)
F test to compare two variances

data:  co$초미세먼지 by co$그룹
F = 0.15931, num df = 30, denom df = 30, p-value = 0.00000267
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.07681741 0.33041032
sample estimates:
ratio of variances
 0.159315
```

#2

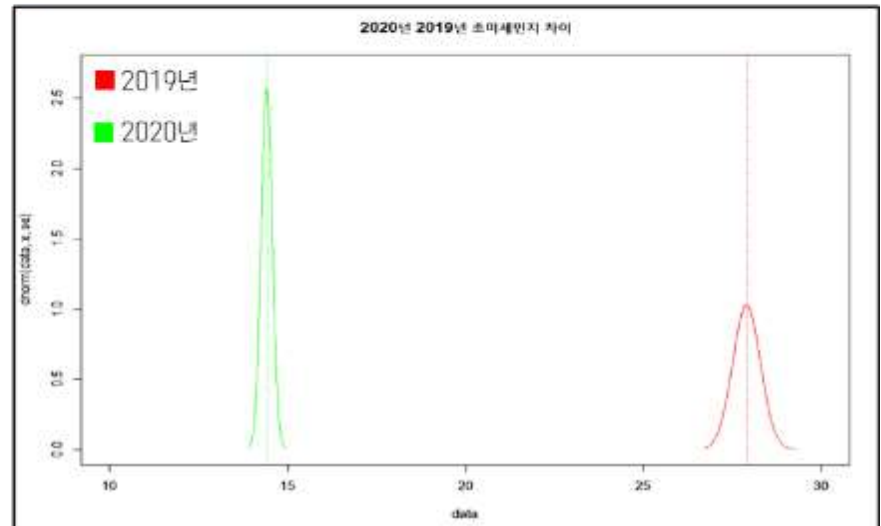
▶ T 검정 (t.test)

```
Welch Two Sample t-test

data:  co$초미세먼지 by co$그룹
t = -4.5598, df = 39.322, p-value = 0.00004898
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -19.463577 -7.504165
sample estimates:
mean in group 2020 mean in group 2019
 14.41935             27.90323
```

#3

▶ 통계 분석 결과 그래프



R STUDIO 분석 2. Paired sample T-Test (대응표본 T 검정)

문제정의

- 코로나 확진자가 증가함에 따라 정부는 사회적 거리두기 2단계를 시행하기로 하였다.
- 사회적 거리두기 2단계 전의 확진자 수와 사회적 거리두기 2단계 후의 확진자 수를 조사하였다.
 - 사회적 거리두기 2단계를 시행한 것과 시행하지 않은 것은 차이가 없다는 귀무가설을 세우고, 이를 검증해보도록 한다.

▶ 데이터 불러오기

```
> b <- read.csv("대응.csv",
+             header=TRUE,
+             na.strings = ".")
> str(b)
'data.frame': 26 obs. of 3 variables:
 $ 지역: chr "강남구" "강동구" "강북구" "강서구" ...
 $ 전 : int 3 2 9 6 6 7 0 3 5 13 ...
 $ 후 : int 7 1 3 1 2 0 1 0 2 3 ...
```

▶ 기본 통계치 확인

```
> describe(b)
vars n mean sd median trimmed mad min max range skew kurtosis
지역+ 1 26 13.50 7.65 13.5 13.50 9.64 1 26 25 0.00 -1.34
전 2 26 5.58 7.37 4.5 4.27 3.71 0 38 38 3.22 11.51
후 3 26 1.35 1.47 1.0 1.14 0.74 0 7 7 2.19 5.81
se
지역+ 1.50
전 1.45
후 0.29
> b2 <- c(b$후-b$전) # 몸무게 차이
> describe(b2)
vars n mean sd median trimmed mad min max range skew kurtosis se
X1 1 26 -4.23 7.46 -3 -3.09 4.45 -37 4 41 -3.16 11.45 1.46
```

▶ 통계분석 (t.test)

```
Paired t-test
data: b$후 and b$전
t = -2.8925, df = 25, p-value = 0.007806
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-7.243198 -1.218341
sample estimates:
mean of the differences
-4.230769
```

#2

해석

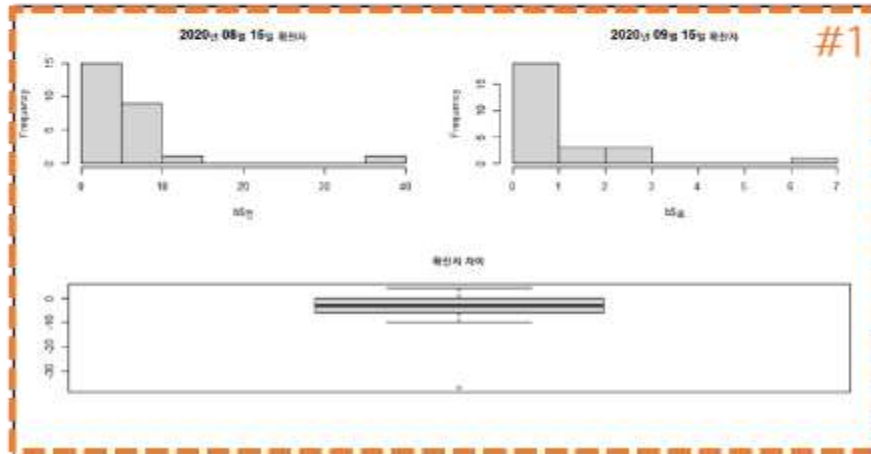
#1 Boxplot과 Histogram을 통해 2020년 8월에 비해 2020년 9월에 확진자 수가 감소하였음을 알 수 있다.

#2 t.test에서 P값이 0.05보다 작기때문에, 귀무가설을 기각하고 2020년 8월과 2020년 9월의 확진자 수는 차이가 있다는 연구가설을 채택하였다.

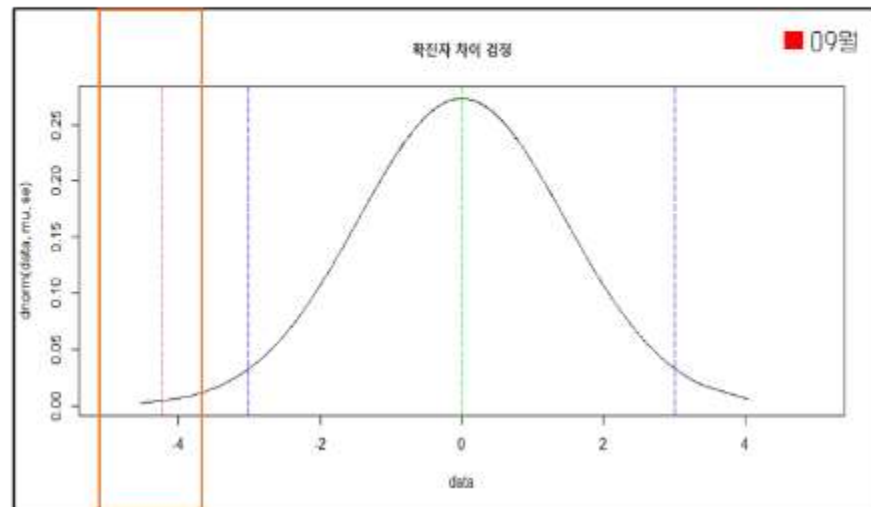
결론

P값이 0.007806이기 때문에 8월과 9월은 통계적으로 유의한 차이가 있으며, 통계 분석 결과 그래프를 보면 평균에서 9월이 벗어나 있기에 2020년 8월과 2020년 9월의 확진자 수 차이가 있다는 사실을 알 수 있다.

▶ Boxplot, Histogram을 통한 데이터 분석



▶ 통계 분석 결과 그래프



- 문제정의**
- 코로나의 전국적 확산으로 인해 대전, 광주, 울산, 제주, 대구의 감염자 발생에 대해 조사하였다.
 - 대전, 광주, 울산, 제주, 대구의 감염자 차이는 없이, 동일하다는 귀무가설을 세우고 이를 검증해보도록 한다.

▶ 데이터 불러오기

```
> # 01. 데이터 불러오기
> a <- read.csv("아노바.csv",
+             header=TRUE,
+             na.strings = ".")
> a$지역 <- factor(a$지역,
+                 levels=c(1:5),
+                 labels=c("대구", "제주", "울산", "광주", "대전"))
> str(a)
'data.frame': 150 obs. of 3 variables:
 $ 지역 : Factor w/ 5 levels "대구", "제주", "...: 1 1 1 1 1 1 1 1 ...
 $ 확진자: int  2 3 0 2 2 0 1 7 6 1 ...
 $ 날짜 : chr  "2020.11.01" "2020.11.02" "2020.11.03"
           "2020.11.04" ...
```

▶ 기본 통계치 확인

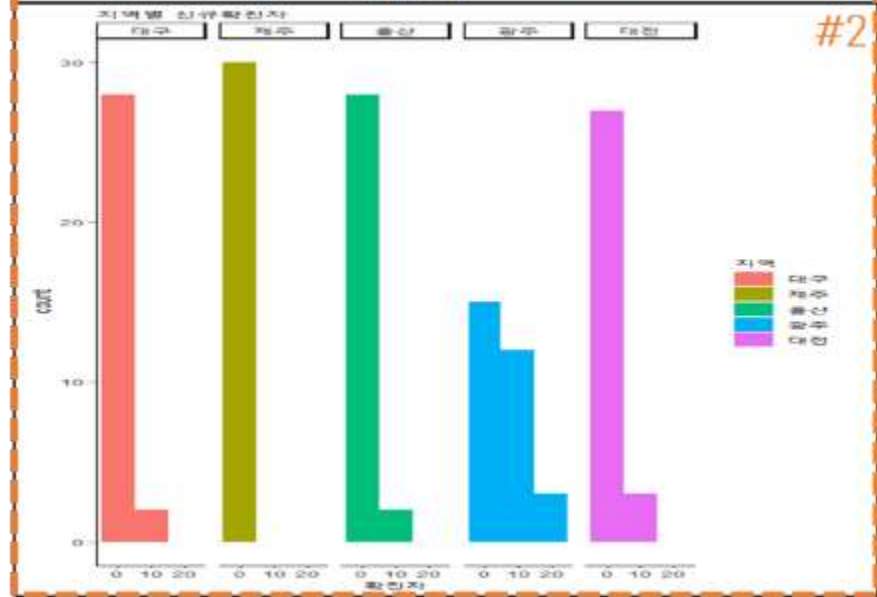
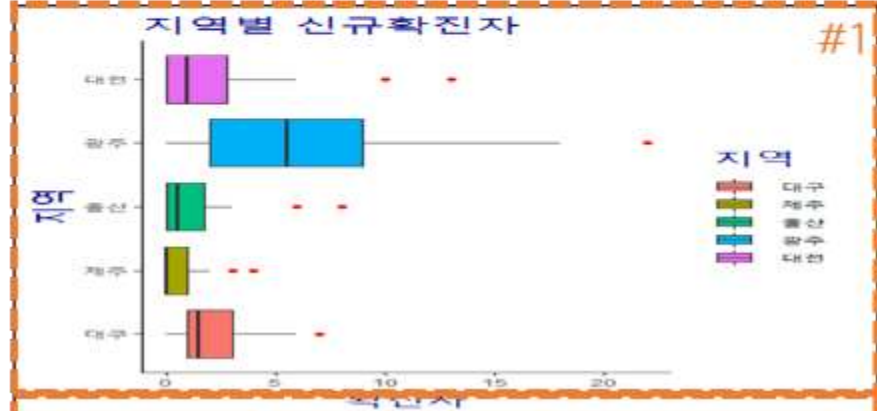
```
> # 02. 기본통계치 확인
> library(psych)
> describeBy(a$확진자, a$지역, mat=T)
  item group1 vars  n   mean   sd median trimmed
x11  1   대구    1 30  1.933333 1.779836  1.5 1.666667
x12  2   제주    1 30  0.700000 1.087547  0.0 0.4583333
x13  3   울산    1 30  1.233333 1.906130  0.5 0.8333333
x14  4   광주    1 30  6.800000 5.950369  5.5 6.1666667
x15  5   대전    1 30  2.066667 3.095417  1.0 1.3750000
      mad min max range   skew  kurtosis   se
x11 1.4826  0  7  7 1.1255510 0.7744945 0.3249521
x12 0.0000  0  4  4 1.5206339 1.3998530 0.1985580
x13 0.7413  0  8  8 1.9871288 3.7851693 0.3480102
x14 5.1891  0 22 22 0.7325849 -0.3651894 1.0863839
x15 1.4826  0 13 13 2.0436887 3.8497427 0.5651433
```

해석

#1 Boxplot을 통해 지역별(대전, 광주, 울산, 제주, 대구)로 신규 확진자 수가 차이가 난다는 사실을 알 수 있다.

#2 Histogram을 통해 제주와 울산은 신규 확진자가 거의 나오지 않았음을 알 수 있고, 광주는 타지역에 비해 신규 확진자가 비교적 많이 나왔음을 알 수 있다.

▶ Boxplot, Histogram을 통한 데이터 분석



▶ 등분산 검정(Bartlett.test)

```
Bartlett test of homogeneity of variances

data: 확진자 by 지역
Bartlett's K-squared = 96.225, df = 4,
p-value < 0.00000000000000022 #3
```

▶ 등분산 검정(LeveneTest)

```
> leveneTest(확진자 ~ 지역,
+           data=a)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F) #3
group  4 14.358 0.0000000006729 ***
      145

---
signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

▶ 이분산 검정

```
> oneway.test(a$확진자 ~ a$지역,
+           data=a,
+           var.equal = FALSE)

One-way analysis of means (not assuming
equal variances)

data: a$확진자 and a$지역
F = 9.7863, num df = 4.000, denom df =
68.917, p-value = 0.000002503 #4
```

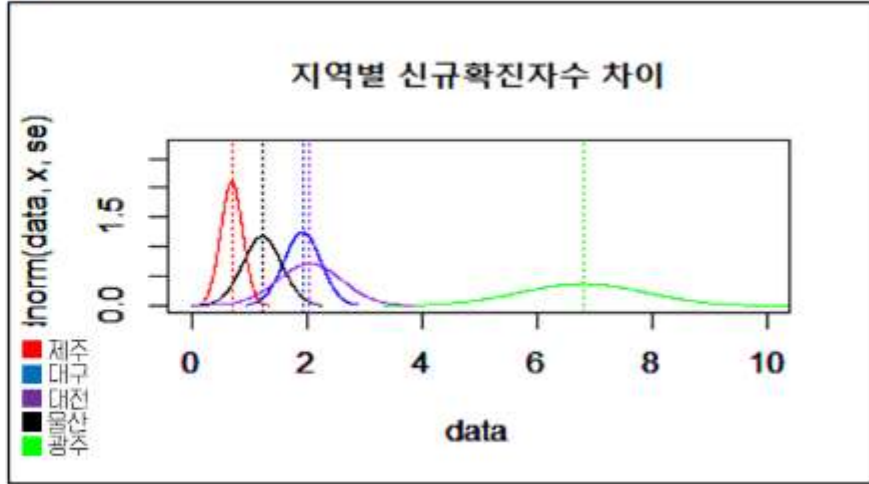
해석

- #3 등분산 검정을 실시한 결과, P값이 0.05보다 작기 때문에 이분산 검정을 임을 알 수 있다.
- #4 이분산 검정을 실시한 결과, P값이 0.05보다 작기 때문에 이는 이분산이 맞고 지역별로 신규 확진자 수 차이가 난다는 연구가설을 채택하였다.
- #5 P값이 0.05보다 작기 때문에, 지역별로 신규 확진자 수에는 차이가 없다는 귀무가설을 기각하고 지역별로 신규 확진자 수에 차이가 있다는 연구가설을 채택하였다.

결론

P값이 0.0000062233829이기 때문에 지역별(대전, 광주, 울산, 제주, 대구)로 신규 확진자 수에 통계적으로 유의한 차이가 있고 광주에서 신규 확진자 수가 가장 많다는 사실을 알 수 있다.

▶ 통계 분석 결과 그래프



▶ library(nparcomp)

```
summary(Result.0)
#> Nonparametric Multiple Comparisons for relative effects
#> Alternative Hypothesis: True differences of relative effects are not equal to 0
#> Estimation Method: Global Pseudo ranks
#> Type of Contrast: Tukey
#> Confidence Level: 95 %
#> Method: Fisher with 51 DF

#> Data Info
#> Sample Size Effect Lower Upper
1  30 0.5420009 0.4825901 0.6044449
2  30 0.3482222 0.2851487 0.3909951
3  30 0.4033333 0.3392448 0.4700156
4  30 0.7251111 0.6569774 0.8000420
5  30 0.4724444 0.4075047 0.5402799

#> Contrast
#> 1 2 3 4 5
2  -1 1 0 0 0
3  -1 0 0 1 0
4  -1 0 0 0 1
5  0 -1 1 0 0
6  0 -1 0 1 0
7  0 0 -1 1 0
8  0 0 1 -1 0
9  0 0 1 0 1
10 0 0 -1 0 1

#> Analysis
#> Estimate Lower Upper statistic p-Value
0  -0.204 -0.952 -0.465 -3.015 0.000004000309
1  -0.143 -0.380 -0.483 -2.287 0.164714804492
2  0.193 0.400 0.389 2.827 0.000022841571
3  0.040 -0.043 0.115 -1.034 0.37145109951
4  0.403 -0.100 0.227 1.962 0.07844482652
5  0.195 0.200 0.335 1.664 0.000000213828
6  0.127 -0.439 0.300 -2.190 0.150159244308
7  0.332 0.181 0.500 2.983 0.000028178458
8  0.075 0.117 0.200 1.700 0.00046054441

#> P-Value
#> Quant 1 2 3 4 5
#> 0.0000062233829 #5
```

R STUDIO 분석 4. Repeated Measured Anova (반복측정 분석)

문제정의

서울시의 코로나 바이러스 확산으로 8월 16일부터 사회적 거리두기 2단계로 격상되었다.
- 사회적 거리두기는 몇 개월을 시행하여도 확진자의 발생에 영향을 주지 않는다는 귀무가설을 세우고 이를 검증해보도록 한다.

▶ 데이터 불러오기

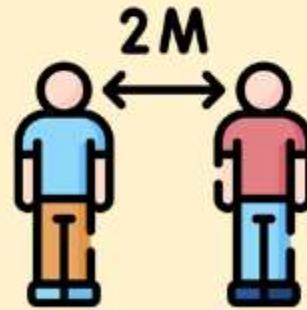
```
# 01. 데이터 불러오기
ccc <- read.csv("covid19.csv",
               header=TRUE,
               na.strings = ".")
str(ccc)
data.frame': 75 obs. of 3 variables:
 $ id : int 1 2 3 4 5 6 7 8 9 10 ...
 $ time : int 1 1 1 1 1 1 1 1 1 1 ...
 $ count: int 11 15 22 35 33 44 24 14 21 14 ...
```

▶ 기본 통계치 확인

```
> describeBy(ccc$count, ccc$time, mat=T)
  item group1 vars n mean sd median
X11 1 격상전 1 25 23.84 10.502698 22
X12 2 1개월 1 25 13.12 5.238638 11
X13 3 2개월 1 25 2.44 3.042477 1
  trimmed mad min max range skew
X11 23.285714 11.8608 10 44 34 0.4056506
X12 12.904762 2.9652 6 23 17 0.6693798
X13 1.952381 1.4826 0 11 11 1.3174914
  kurtosis se
X11 -1.0434645 2.1005396
X12 -1.1109783 1.0477277
X13 0.8276516 0.6084954
```

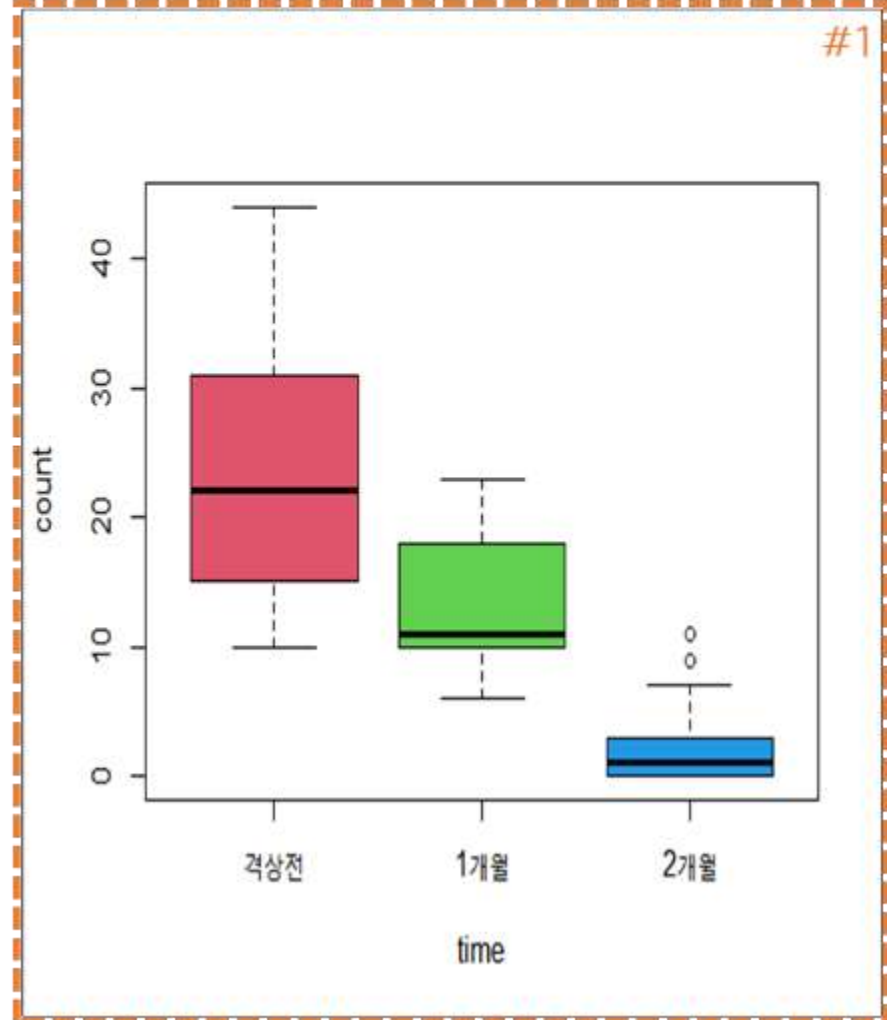
해석

#1 Boxplot을 통해 사회적 거리두기 격상 전, 2단계 격상 후 1개월 후, 2단계 격상 후 2개월 후 확진자 추이가 감소하였음을 알 수 있다.



사회적거리두기

▶ Boxplot을 통한 데이터 분석



R STUDIO 분석 4. Repeated Measured Anova (반복측정 분석)

▶ 구형성 검정(Mauchly's Test of Sphericity)

```
Univariate Type III Repeated-Measures ANOVA Assuming Sphericity
```

	Sum Sq	num Df	Error SS	den Df	F value	Pr(>F)
(Intercept)	12936.3	1	1160.7	24	267.495	0.00000000000001627
time.f	5724.5	2	2367.5	48	58.031	0.00000000000015474

```
(Intercept) ***
time.f      ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Mauchly Tests for Sphericity
```

	Test statistic	p-value
time.f	0.62259	0.0042987

해석

- #2 각 반복측정치들간의 상관관계가 모두 동일한지에 대해 알아보고, 구형성 검정을 실시한 결과 P값이 0.05보다 작기때문에 구형성 조건을 만족하지 않는다.
- #3 time.f의 P값 또한 0.05보다 작기때문에 격상 전, 2단계 격상 후 1개월 후, 2단계 격상 후 2개월 후의 확진자 수 차이가 있다는 사실로서 옳은 검정임을 알 수 있다.
- #4 일변량 Anova 검정에서 P값이 0.05보다 작기때문에 사회적거리두기는 확진자 발생에 영향을 준다는 사실을 알 수 있다.

▶ 사후 검정(TukeyHSD)

```
Simultaneous Tests for General Linear Hypotheses
```

Multiple Comparisons of Means: Tukey Contrasts

```
Fit: lm(formula = count ~ time, data = ccc)
```

```
Linear Hypotheses:
```

	Estimate	Std. Error	t value	Pr(> t)
1개월 - 격상전 == 0	-10.72	1.98	-5.414	<0.00001 ***
2개월 - 격상전 == 0	-21.40	1.98	-10.808	<0.00001 ***
2개월 - 1개월 == 0	-10.68	1.98	-5.394	<0.00001 ***

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)
```

▶ 일변량 ANOVA 분석

```
Error: id
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Residuals	1	92.31	92.31		

```
Error: id:time
```

	Df	Sum Sq	Mean Sq
time	2	4163	2081

```
Error: Within
```

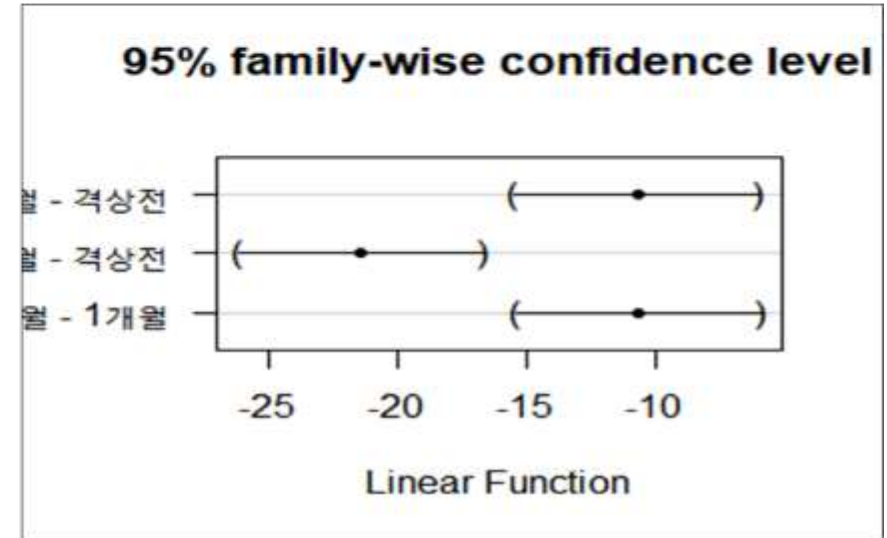
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
time	2	1584	792.2	16.01	0.00000194 ***
Residuals	69	3413	49.5		

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

결론

사후검정의 P값이 전부 0.05보다 작다는 사실을 통해 격상 전, 2단계 격상 후 1개월 후, 2단계 격상 후 2개월 후의 확진자 수 차이가 없다는 귀무가설을 기각하고, 차이가 있다는 연구가설을 채택하였다. 따라서 사회적거리두기는 확진자 수 차이에 통계적으로 유의미한 영향을 준다는 사실을 알 수 있기때문에 10월 12일에 1단계가 되었다.

▶ 통계 분석 결과 그래프



R STUDIO 분석 5. 단순 회귀분석

문제정의 연령별 코로나 확진자와 사망자의 추이는 상이하게 나타난다. 연령에 따른 확진자 수는 사망자 수에 영향을 미치지 않는다는 귀무가설을 세우고 이를 검증해보도록 한다.

▶ 데이터 불러오기

```
death <- read.csv("death.csv",
                  header=TRUE,
                  na.strings = ".")
```

▶ 기본 통계치 확인

```
> describe(death)
```

	vars	n	mean	sd	median	trimmed	mad	min	max
연령	1	9	40.00	27.39	40	40.00	29.65	0	80
확진자	2	9	5049.11	2592.03	5790	5049.11	3313.61	1462	8315
사망자	3	9	68.00	111.72	6	68.00	8.90	0	315
	range	skew	kurtosis	se					
연령	80	0.00	-1.60	9.13					
확진자	6853	-0.10	-1.84	864.01					
사망자	315	1.22	-0.11	37.24					

▶ Anova 분석(1)

```
> anova(death.model)
```

Analysis of Variance Table

Response: 사망자

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
연령	1	66733	66733	14.106	0.007114 **
Residuals	7	33117	4731		

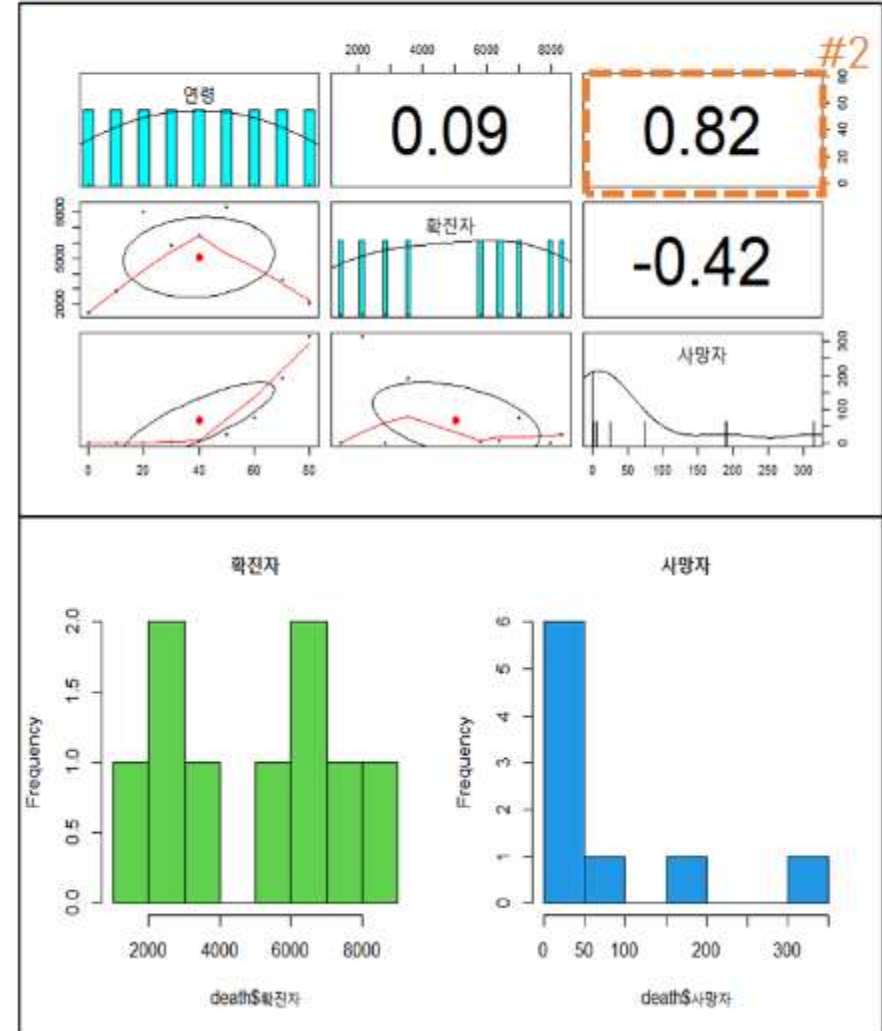
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

해석

#1 연령별 코로나 확진자와 사망자 추이의 상관관계수인 P값이 0.05 미만이므로, 통계적으로 유의한 양의 관계가 있는 것을 알 수 있다.

#2 상관관계수가 연령자별 사망자가 0.82로써 밀접한 연관성이 있다고 할 수 있다. 하지만, 확진자와 사망자의 연관성은 -0.42로 없다고 볼 수 있다.

▶ 산점행렬도, Histogram을 통한 데이터 분석



▶ Anova 분석(2)

```
> summary(death.model)

Call:
glm(formula = 사망자 ~ 연령, data = death)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-76.35  -60.70  -1.30   32.05  113.60

Coefficients:
(Intercept) -65.400      42.276     -1.547     0.16579
연령         3.335       0.888       3.756     0.00711 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 4730.95)

Null deviance: 99850  on 8  degrees of freedom
Residual deviance: 33117  on 7  degrees of freedom
AIC: 105.44

Number of Fisher Scoring iterations: 2
```

#3

▶ 잔차의 등분산성 검정

```
> # 잔차의 등분산성 검정
> car::ncvTest(death.model)
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 1.246555, Df = 1, p = 0.26421
```

#4

▶ 잔차의 정규분포 검정

```
> # 잔차의 정규분포 검정
> shapiro.test(death.model$residuals)

Shapiro-Wilk normality test
data:  death.model$residuals
W = 0.94285, p-value = 0.6122
```

#4

해석

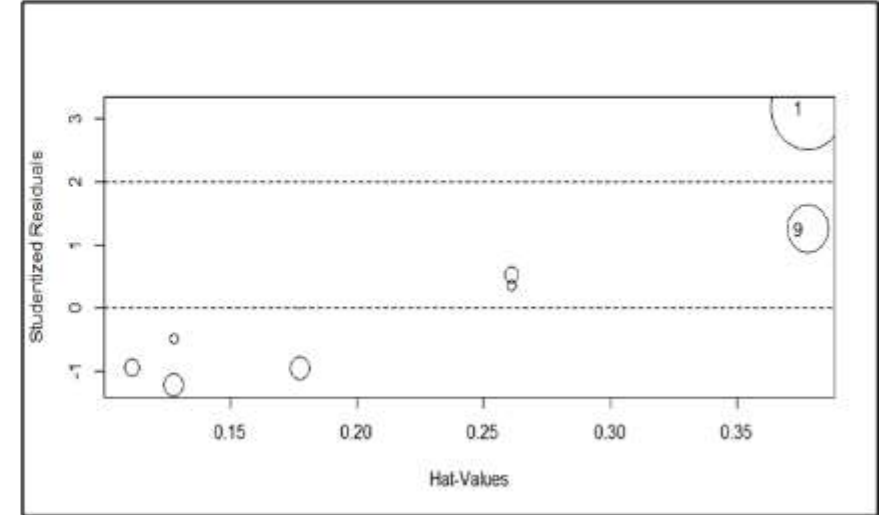
#3 ANOVA검정 결과 P값이 0.05보다 작으므로, 회귀선은 통계적으로 유의하게 영향력이 있다.

#4 잔차의 등분산성 검정과 잔차의 정규분포 검정의 P값이 0.05보다 크므로 이는 정규성을 띤다.

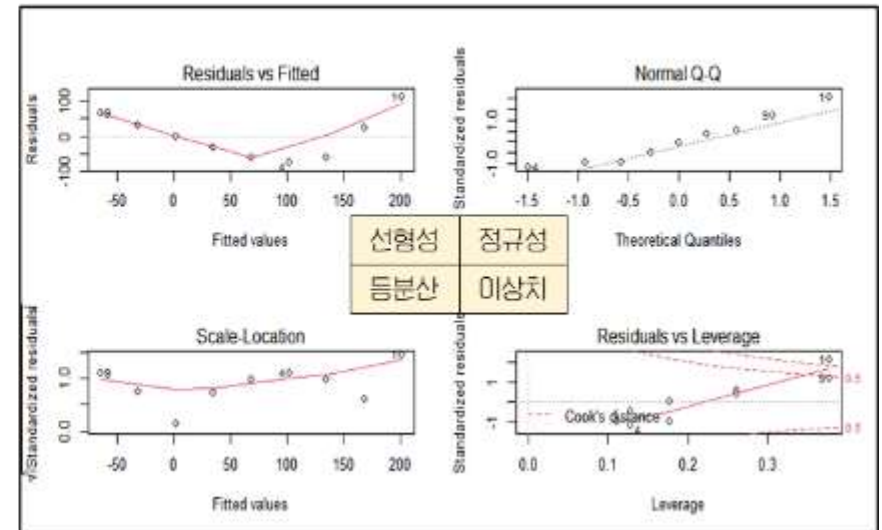
결론

ANOVA검정 결과 회귀선은 통계적으로 유의하게 영향력이 있으며 (P=0.00711), 회귀식은 $y=3.335x-65.400$ 으로 나타났다. 연령은 연령에 따른 확진자 사망률의 결과에 유의미하며 두 사이의 관계를 설명할 수 있는 설명력이 82%로 측정되었다. 결국 연령은 사망률에 영향을 미치고, 확진자는 사망률에 영향을 크게 미치지 않는다는 것을 확인할 수 있다.

▶ 이상치 검정



▶ 가정 검정

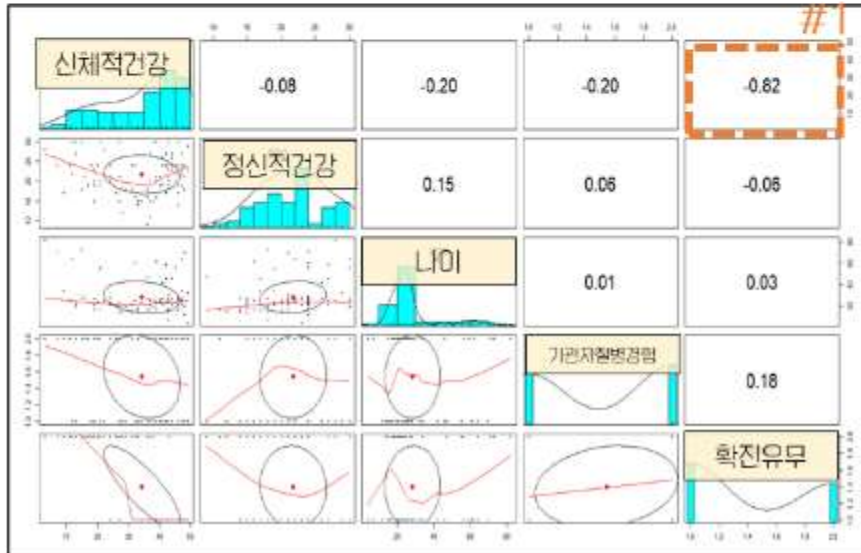


문제정의 2팀 조장 김○○씨는 코로나 확진에 “신체적 건강점수와 심리적 건강점수, 나이, 기관지질환을 앓았던 경험”이 영향을 준다고 보고, 이들 간의 인과관계를 확인해보고자 한다. (신체적 건강점수와 심리적 건강점수는 50점 만점이다)

▶ 데이터 불러오기 및 기본통계치 확인

```
> # 01. 데이터 불러오기
> lo <- read.csv("로지스틱.csv",
+               header=TRUE,
+               na.strings = ".")
> describe(lo)
  vars  n  mean  sd median trimmed  mad min max range  skew kurtosis
phy    1 100 34.30 12.26   39  35.56 10.38   3  49  46 -0.72   -0.78
psy    2 100 21.78  4.88   22  21.80  4.45   9  30  21 -0.12   -0.56
age    3 100 28.44 15.13   24  25.65  5.19   4  82  78  1.72    2.34
bron*  4 100  1.54  0.50    2   1.55  0.00    1  2   1 -0.16   -1.99
chun*  5 100  1.40  0.49    1   1.38  0.00    1  2   1  0.40   -1.86
  se
phy  1.23
psy  0.49
age  1.51
bron* 0.05
chun* 0.05
```

▶ 산점행렬도를 통한 데이터 분석



해석

- #1 코로나 확진은 신체적 건강점수와 가장 밀접한 상관관계가 있음을 알 수 있다(상관관계가 -0.82로 음수 이므로 반비례 관계임이 확인된다)
- #2 로지스틱 회귀분석 결과 유의확률 (P<0.05)이 신체건강점수, 정신적건강점수의 변수가 코로나 확진에 유의미한 영향을 준다.
- #3 신체적건강점수가 높아지면, 코로나 확진은 0.7배 내려간다. 또한 기관지 질환을 앓았던 경험이 있는 사람은 기관지 질환을 앓았던 경험이 없는 사람보다 코로나 확진이 될 확률이 1.9배 높아진다.

결론

김○○씨는 확진률을 낮추기 위해 신체적 건강점수와 기관지 질환 경험에 집중해야 할 필요가 있다. 이로서 코로나 확진을 낮추기 위해서는 신체적 건강에 만전을 기울이고, 기관지 질환에 걸린 경험이 있는 사람들은 특별히 주의해야 한다.

▶ 로지스틱 회귀분석

```
> summary(lo)
Call:
glm(formula = chun ~ phy + psy + age + bron, family = binomial,
    data = lo)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.6361  -0.2510  -0.1065   0.1301   2.5218

Coefficients:
(Intercept) 13.95388 3.30790 4.2185 0.00002461 ***
phy         -0.33091 0.06799 -4.867 0.00000113 ***
psy         -0.11960 0.07966 -1.501 0.133
age         -0.03708 0.03605 -1.029 0.304
bronYes     0.67598 0.86870 0.778 0.436

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

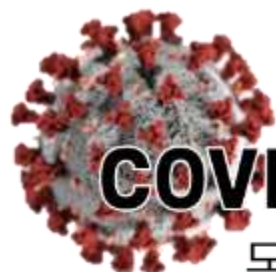
(Dispersion parameter for binomial family taken to be 1)

Null deviance: 134.602 on 99 degrees of freedom
Residual deviance: 39.384 on 95 degrees of freedom
AIC: 49.384

Number of Fisher Scoring iterations: 7
```

▶ Odds 계산

```
> # Odds 계산
> odds <- data.frame(summary(lo)$coefficients,
+                   odds = exp(coef(lo)))
> round(odds, 5)
  Estimate Std. Error z.value Pr(>|z|) odds
(Intercept) 13.95388 3.30790 4.21835 0.00002 1148403.15361
phy         -0.33091 0.06799 -4.86684 0.00000 0.71827
psy         -0.11960 0.07966 -1.50131 0.13328 0.88727
age         -0.03708 0.03605 -1.02858 0.30367 0.96359
bronYes     0.67598 0.86870 0.77815 0.43648 1.96596
```



COVID19

목적 : 코로나가 현대사회에 시사하는 문제점

분석결과

- 얻은 것 : 초미세먼지농도 감소,
지역별로 확진자 수 상이,
사회적 거리두기 단계에 따라 확진자 수 감소
- 활용가능기관 : 공공기관(질본, 시청), 기업(마케팅 전략 수립 가능)

감사합니다