

Covid-19 우세종에 따른 감염자 수 정설 검증

진화를 거듭할수록 감염성이 높아지고 치명성이 줄어드는가?

60211642 김광영 60211670 송지은 60211673 신유빈



01

문제 정의

기획의도, 문제 정의, 단어 정의

•

02

데이터 분석

자료 수집, 분석 순서 정의, 분석 코드

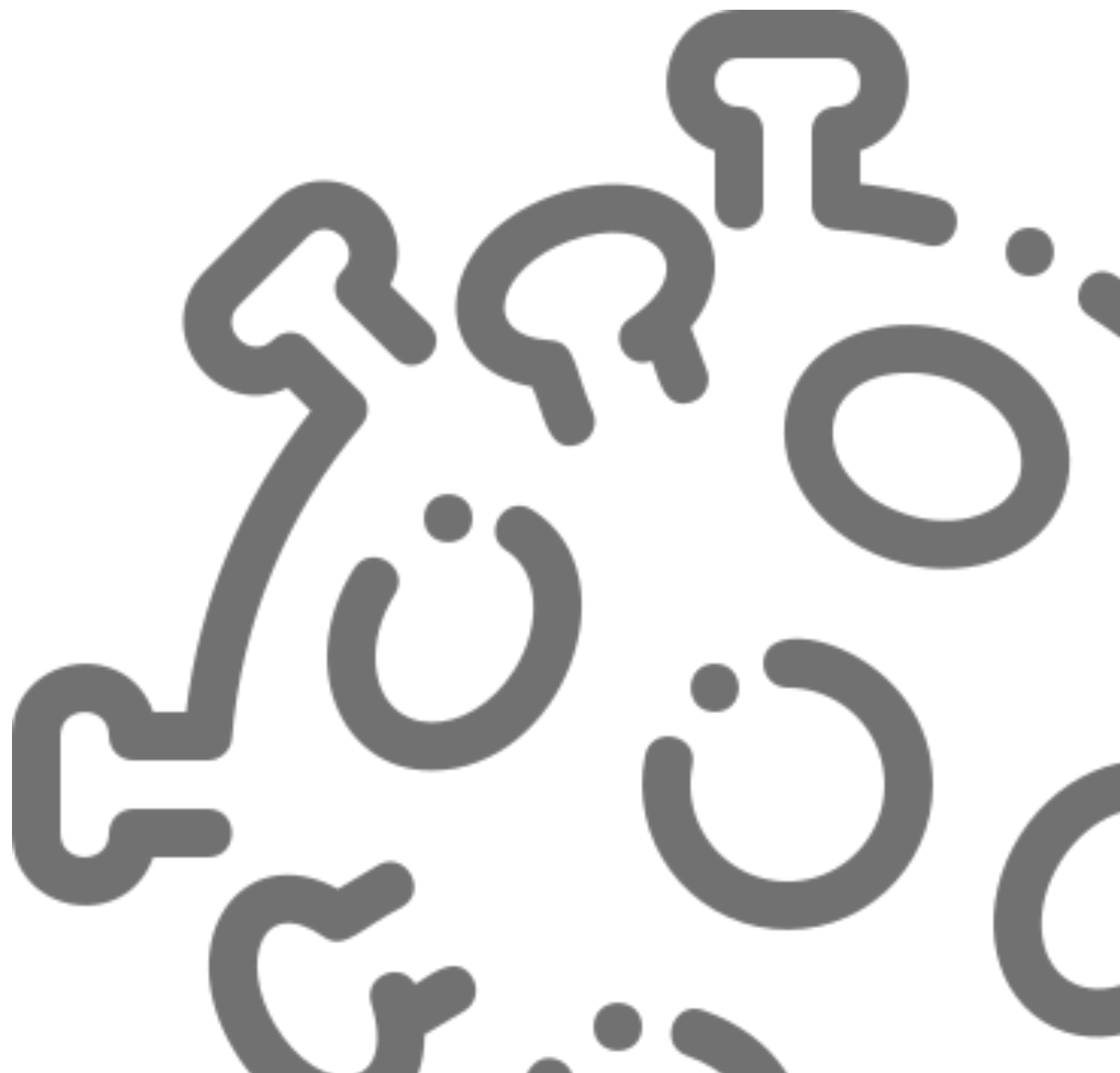
•

03



결과



얻은 결론, 아쉬운 점



문제정의





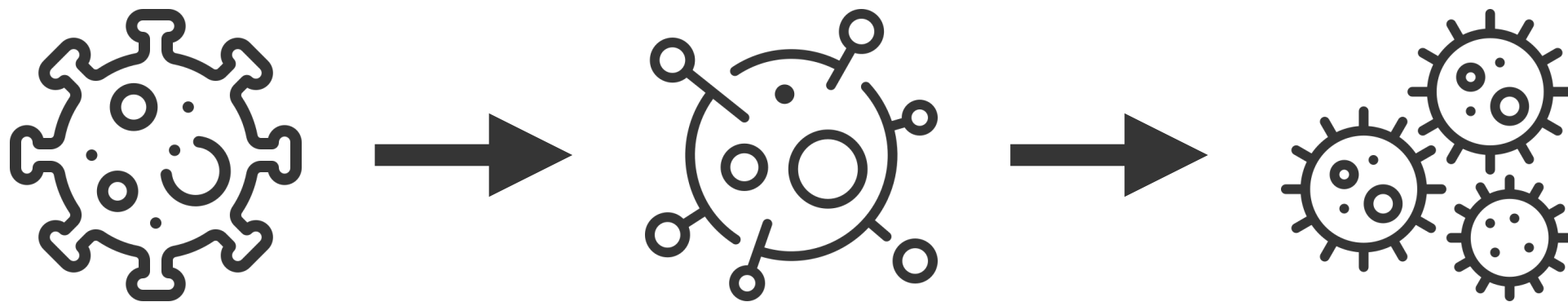
 조선일보 |  A12면 1단 | 2022.11.27. | 네이버뉴스
코로나 중증화·사망률, 델타변이 때의 20분의 1 수준
올 하반기 오미크론 하위 변이인 BA.5가 국내 **우세종**이 된 뒤 **코로나**

 한겨레 |  9면 1단 | 2022.07.05. | 네이버뉴스
전파속도 빠른 'BA.5' 변이 우세종 우려...코로나 재확산 한다
5)가 조만간 **우세종**이 될 전망이다 가운데, 방역당국이 예상치를 뛰어넘는 재확산

 조선일보 |  A12면 1단 | 2022.11.16. | 네이버뉴스
하루 접종자 10만명 근접... 우세종 대응 개량 백신에 몰렸다
BA.4·5 백신 접종 56% 달해 현재 국내 **코로나** 유행을 주도하는 오미크론 변이



**바이러스가 진화할수록
치명률은 낮아지고 감염률은 높아지게 진화한다?**



- 증명할 가정
 - 바이러스가 진화할수록 치명률은 낮아지고 감염률은 높아지게 진화한다
- 현재 한국에서 보인 코로나 바이러스의 우세종은 "델타", "오미크론", "그 외"로 분류한다.
- WHO에서 추적한 변종 표를 참고하여 분류기준을 정하였다.

WHO label	Pango lineage*	GISAI clade	Nextstrain clade	Earliest documented samples	Date of des
Alpha	B.1.1.7	GRY	20I (V1)	United Kingdom, Sep-2020	VOC: 18-Dec-2020 Previous VC 2022
Beta	B.1.351	GH/501Y.V2	20H (V2)	South Africa, May-2020	VOC: 18-Dec-2020 Previous VC 2022
Gamma	P.1	GR/501Y.V3	20J (V3)	Brazil, Nov-2020	VOC: 11-Jan-2021 Previous VC 2022
Delta	B.1.617.2	G/478K.V1	21A, 21I, 21J	India, Oct-2020	VOI: 4-Apr-2021 VOC: 11-May-2021 Previous VC 2022

WHO에서 추적한 변종 표
<https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/>



- 귀무가설(H0) — 코로나 바이러스가 진화를 거듭해도 감염률과 치명률은 변화가 없다.
- 대립가설(H1) — 코로나 바이러스가 진화를 거듭하면 감염률과 치명률에 변화가 생긴다.

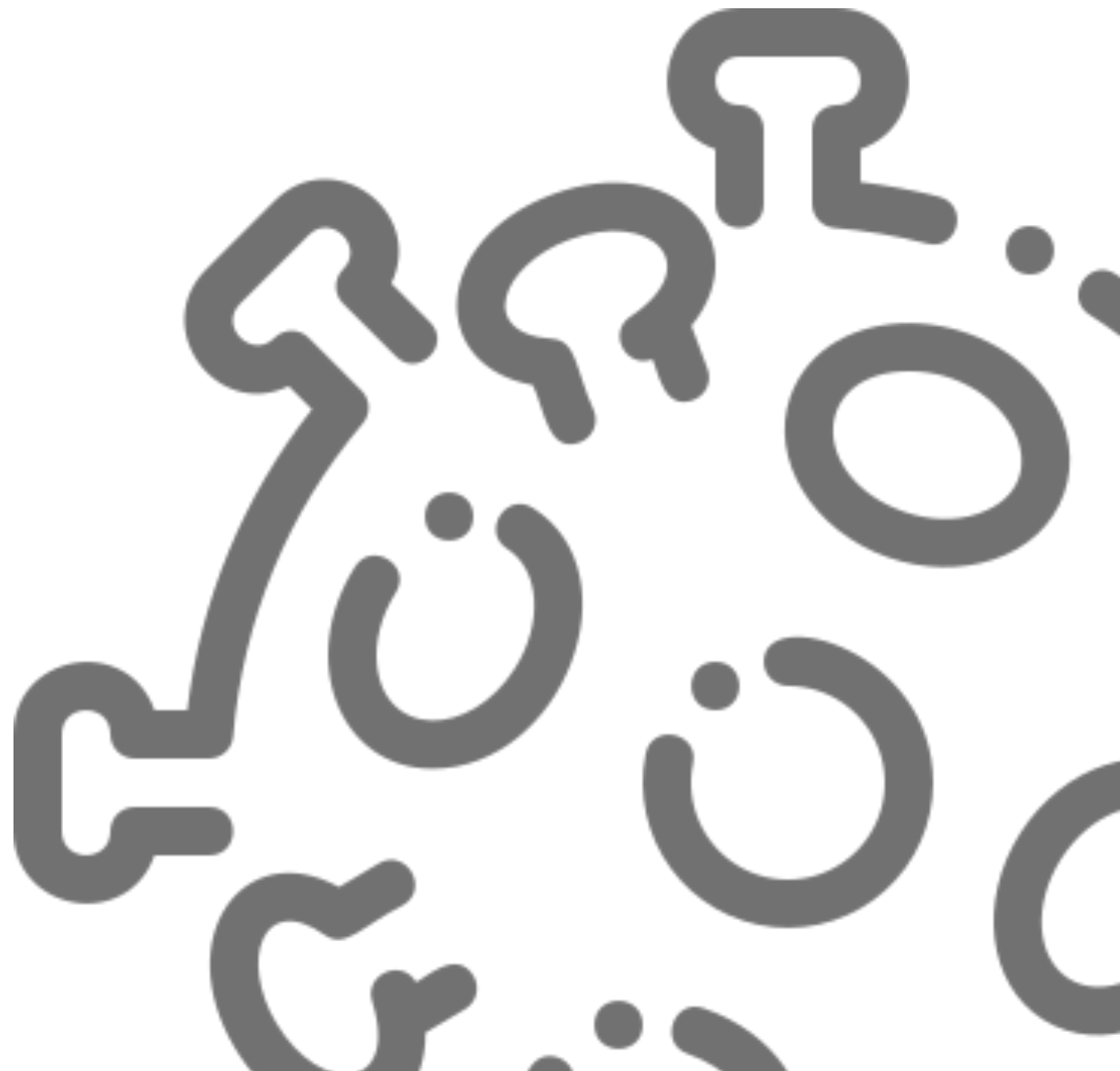


$$p_i = \frac{\text{1일감염자수}}{\text{전체국민수} - \text{전날까지의누적감염자수}}$$

$$p_l = \frac{\text{1일사망자수} \times 20}{\text{17일이전의누적확진자수}}$$

- 우세종은 국내 검출율(국내 + 해외유입사례)이 50%를 넘을 때를 기준으로 파악한다.
- 감염률(p_i)은 1일 감염자 수 / (전체국민 수 - 전날까지 누적 감염자 수)로 계산한다.
(i는 infection의 약자)
- 치명률(p_l)은 1일 사망자 수*20 / 17일 이전의 누적 확진자 수로 계산한다.
(l은 lethality의 약자) (20을 곱한 것은 한겨레 기사의 치명률 정의를 따랐다.)
- 사망자의 사망원인이 사망자로 집계된 당일의 우세종에 영향을 받지 않았을 것이다.
따라서 17일 이전의 확진자에서 사망자가 나올 확률이 높다.

데이터 분석



- 자료 수집



- 확진자와 사망자에 대한 데이터는 한국의 질병관리청 홈페이지를 이용하였다.

- 기간별 우세종을 기준으로 데이터 나누기

우세종	우세종이 된 시점
오미크론	22년 1월 24일
델타	21년 7월 25~31일

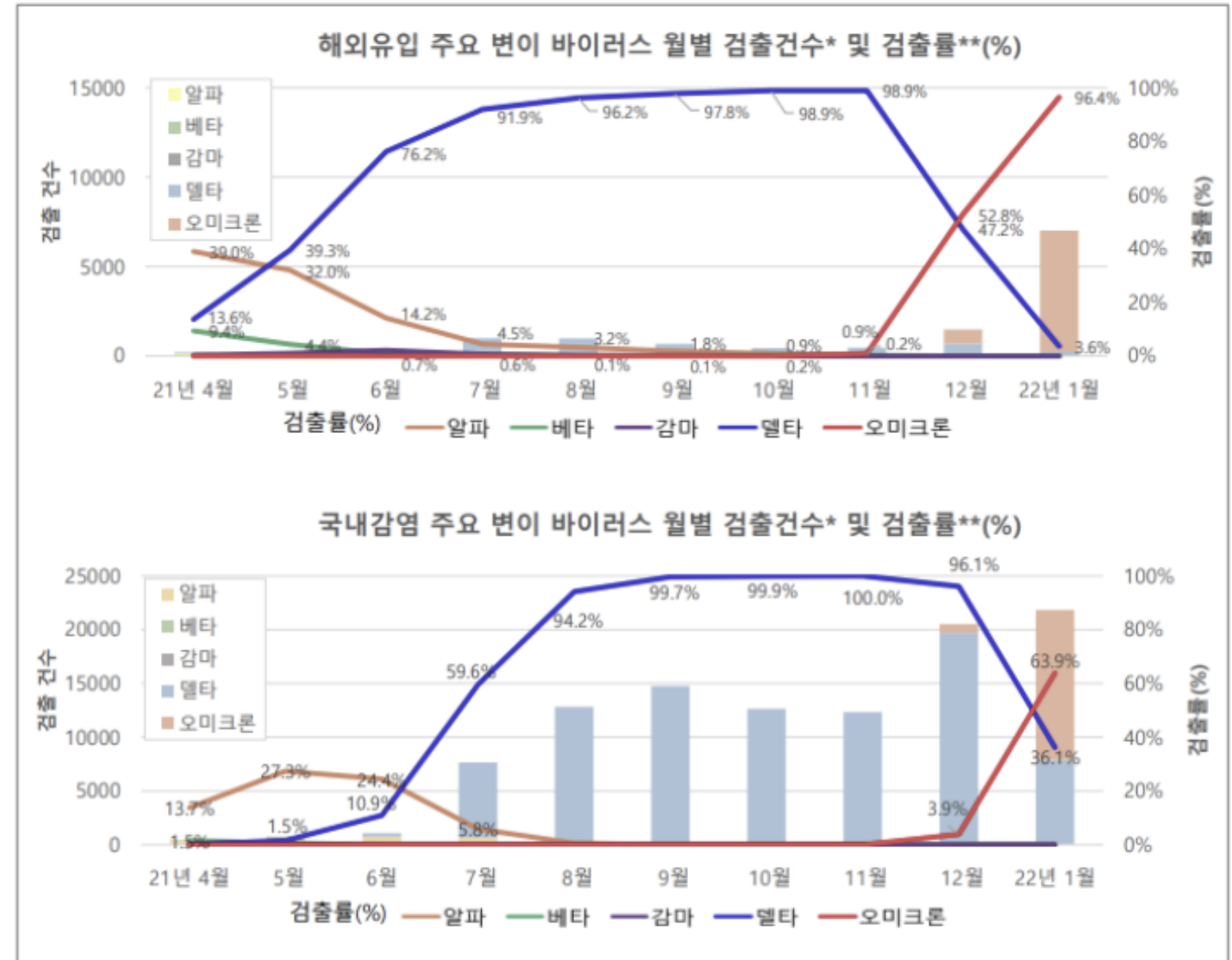


그림 3. 주요 변이 바이러스(VOC) 확인 건수 및 월별 검출률

*확인일 기준. **변이 바이러스 검출률(%) = (변이 바이러스 수 / 분석 건수) × 100

주간 건강과 질병 제 15권 제 8호 p.497~p.510 2022년 1월 국내 코로나19 변이 바이러스 발생 현황 및 특성
중앙방역대책본부 진단분석단 검사분석팀 김일환, 박애경, 이혁진, 김희만, 김준영, 김정아, 노진선, 이채영, 이지은, 김은진*

<https://www.bbc.com/korean/news-59990526>

https://www.voakorea.com/a/korea_korea-social-issues_korea-coronavirus-7/6060704.html



**3개의 집단으로 분류되기 때문에 ANOVA test를 통해 집단 간 평균의 차이를 입증하고,
이후 각각 2개 집단에 대한 비교를 진행해 감염률의 상승과 치명률의 감소를 볼 것이다.**

진행과정

1.데이터를 날짜를 기준으로 3가지 그룹으로 분류했다.
(오미크론, 델타, 그 외)

```
df_covid_etc = df_covid_rmOutlier.loc['2020.4.2':"2021.7.31"]  
df_covid_delta = df_covid_rmOutlier.loc["2021.8.1":"2022.1.24"]  
df_covid_omi = df_covid_rmOutlier.loc["2022.1.25":]
```

진행과정

2. 감염률, 치명률을 아래 정의에 따라 구하고, 신천지 집단 감염 사태의 데이터를 outlier 처리해 데이터에서 제외했다.

$$p_i = \frac{\text{1일감염자수}}{\text{전체국민수} - \text{전날까지의누적감염자수}}$$

$$p_l = \frac{\text{1일사망자수} \times 20}{\text{17일이전의누적확진자수}}$$

#감염률 구하는 코드

```
df_covid["감염률"] = df_covid["계(명)"]
def infection_rate(i):
    rok_pop = 51628117
    if i == 0:
        df_covid["감염률"][0] = int(df_covid["계(명)"][0])/rok_pop
    else:
        denominator = (rok_pop - df_covid["누적 확진자 수"][i-1])
        numerator = df_covid["계(명)"][i]
        p_infection = numerator/denominator
        df_covid["감염률"][i] = p_infection

for i in range(len(df_covid)):
    infection_rate(i)
```

#치명률 구하는 코드

```
df_covid["치명률"] = df_covid["누적 사망자 수"]
def death_rate(i):
    if i < 17:
        df_covid["치명률"][0] = 0
    else:
        p_lethality = int(df_covid["누적 사망자 수"][i])/int(df_covid["누적 확진자 수"][i-17])
        df_covid["치명률"][i] = p_lethality

for i in range(len(df_covid)):
    death_rate(i)
```

신천지 집단 감염 사태와 코로나 유행 초기인 걸 생각했을 때 앞부분 데이터가 신뢰도가 떨어질 것.
신천지 사태 한 달 후인 3월 17일 이후 데이터만 사용하는게 치명률에서 이상치를 줄일 수 있을 것이라 생각.
df_covid_rmOutlier = df_covid.loc['2020.3.17':]

진행과정

3. 정규성 검증인 Shapiro test와 등분산 검증인 fligner test를 진행했다.

[감염률 정규성 검증]

Shapiro Test-statistics : 0.8539105653762817, p-value : 7.790196369742615e-21

Shapiro Test-statistics : 0.835544228553772, p-value : 7.48299046424844e-13

Shapiro Test-statistics : 0.805178165435791, p-value : 7.125169911372918e-19

[감염률 등분산 검증]

Fligner Test-statistics : 683.7901745808464, p-value : 3.2873822050209747e-149

[치명률 정규성 검증]

Shapiro Test-statistics : 0.9608435034751892, p-value : 4.450002033529188e-10

Shapiro Test-statistics : 0.9434995055198669, p-value : 1.8121364746548352e-06

Shapiro Test-statistics : 0.5536781549453735, p-value : 3.1924320684309763e-27

[치명률 등분산 검증]

Fligner Test-statistics : 148.9384212608489, p-value : 4.554416576790575e-33

진행과정

4. 세 집단 평균의 차이를 증명해야 하기 때문에 Kruskal-Wallis Test를 진행해 차이를 밝혔다.

[감염률 Kruskal-Wallis Test]

Kruskal-Wallis Test-statistics : 804.3463264524516, p-value : 2.1797889228493447e-175

[치명률 Kruskal-Wallis Test]

Kruskal-Wallis Test-statistics : 798.2136625406547, p-value : 4.678487671713772e-174

진행과정

4. 마지막으로 각각의 군끼리 rank-sum test 를 실행한다.

```
# Wilcoxon rank-sum test 실행

# df_covid_etc
# df_covid_delta
# df_covid_omi

print("감염률")
test_stat, p_val = stats.ranksums(df_covid_omi['감염률'], df_covid_delta['감염률'])
print("rank-sum Test-statistics omi-delta : {}, p-value : {}".format(test_stat, p_val))
test_stat, p_val = stats.ranksums(df_covid_etc['감염률'], df_covid_delta['감염률'])
print("rank-sum Test-statistics etc-delta: {}, p-value : {}".format(test_stat, p_val))
test_stat, p_val = stats.ranksums(df_covid_omi['감염률'], df_covid_etc['감염률'])
print("rank-sum Test-statistics omi-ect: {}, p-value : {}".format(test_stat, p_val))

print("치명률")
test_stat, p_val = stats.ranksums(df_covid_omi['치명률'], df_covid_delta['치명률'])
print("rank-sum Test-statistics omi-delta : {}, p-value : {}".format(test_stat, p_val))
test_stat, p_val = stats.ranksums(df_covid_etc['치명률'], df_covid_delta['치명률'])
print("rank-sum Test-statistics etc-delta: {}, p-value : {}".format(test_stat, p_val))
test_stat, p_val = stats.ranksums(df_covid_omi['치명률'], df_covid_etc['치명률'])
print("rank-sum Test-statistics omi-ect: {}, p-value : {}".format(test_stat, p_val))

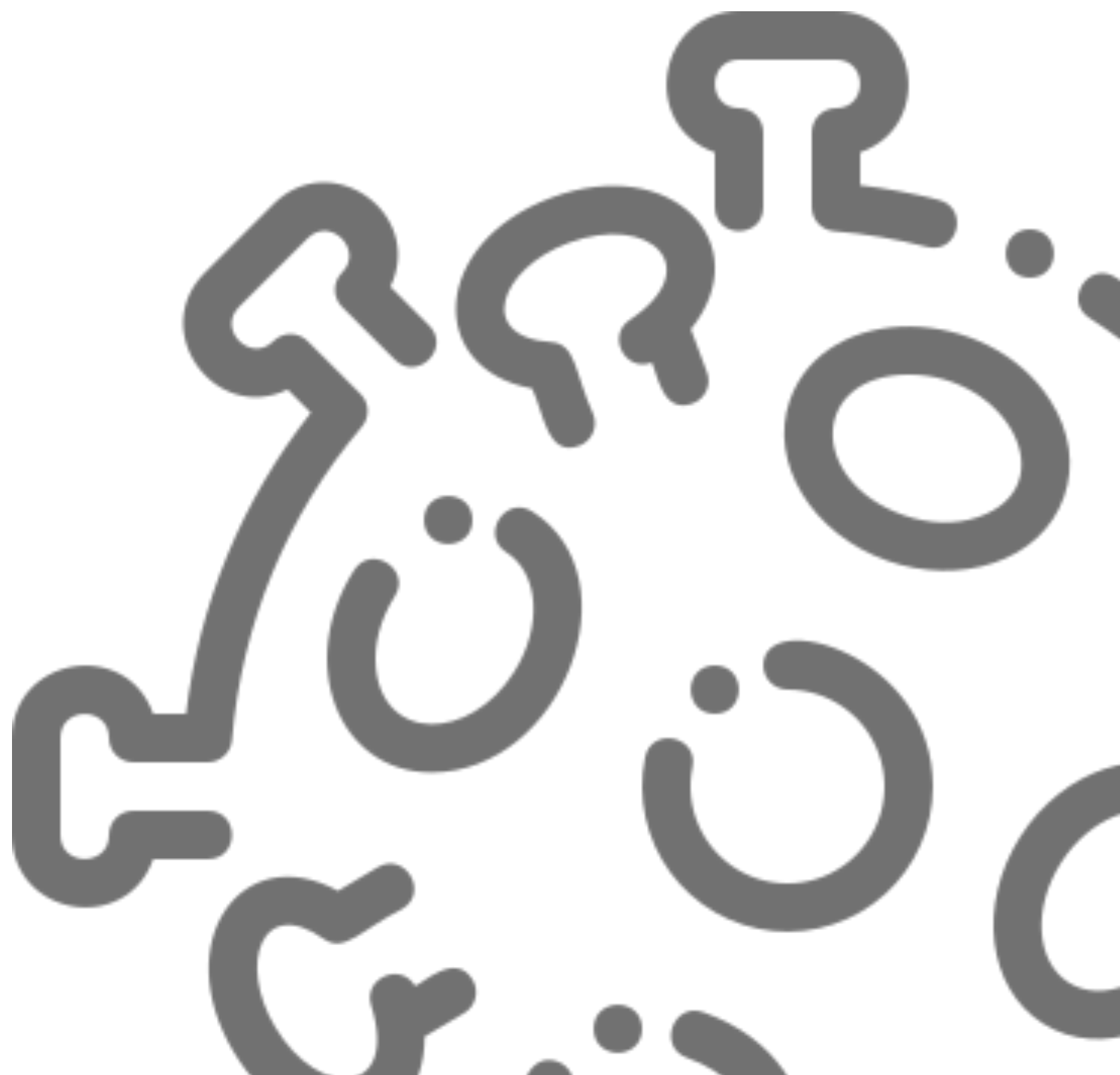
감염률
rank-sum Test-statistics omi-delta : 17.552398698639596, p-value : 5.702014276527441e-69
rank-sum Test-statistics etc-delta: -19.15796396197209, p-value : 8.305931597068906e-82
rank-sum Test-statistics omi-ect: 23.91238424208976, p-value : 2.276712518052433e-126
치명률
rank-sum Test-statistics omi-delta : -17.272640142245073, p-value : 7.560317451411355e-67
rank-sum Test-statistics etc-delta: 17.197760405441255, p-value : 2.759946139893418e-66
rank-sum Test-statistics omi-ect: -21.255452659635345, p-value : 2.935181598757898e-100
```

각각 군들의 p-value 가 모두 유의수준 0.05 보다 작으므로 귀무가설을 기각한다.

따라서 각 군들의 평균은 다르다는 결론을 얻을 수 있다.

bartlett test 를 통해 유의수준을 0.05/3 까지 낮추어도 귀무가설을 기각 할 수 있다.

결과



문제점들

배운 내용의
한계

시기에 따른
각 변이의 비율
미적용

데이터 특성에 따른
outlier의 정의와 처리의 미숙

