

**한국의 코로나19 추이 분석을 통한**

**2023년 기준 한국의 코로나 정책 타당성 파악**

**- CSSE 데이터셋을 바탕으로**

탐색적 데이터 구축

2023.03.15

**B1팀**

**김예지, 서영석, 이현빈, 전국림**

**<목차>**

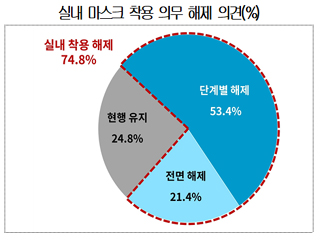
1. **서론**
2. 데이터 분석 배경 p. 3
3. 데이터 분석 설명 p. 3
4. **본론**
5. 한국의 코로나 환자 추세 분석 p. 4
6. 한국의 코로나 환자 통계 p. 5
7. 한국의 월별 코로나 환자 분석 p. 7
8. **분석 결과 및 결론**
9. 데이터 분석 결과 및 결론 도출 p.10

**참고 자료**

**1. 서론**

**1) 데이터 분석 배경**

2023년 현재, 지난 2020년부터 세계를 강타한 코로나 19 팬더믹이 위드 코로나(With Covid19) 시대로 완전히 넘어가고 있다. 이러한 세계적 추세에 맞추어 한국도 지난 2022년 하반기부터 위드코로나 정책을 실시하고 있다. 실내 마스크 의무 착용 해제, 시설 방문객 열 측정 의무 해지, 코로나 지정 병원 축소, 외국인 입국자 무격리 정책 등이 대표적인 예로써 존재한다.



[표1] 실내마스크 착용의무 해제애 대한 의견 (데일리팜)

그러나 국내에는 아직까지 코로나19에 대한 보수적인 의견이 팽배하게 자리매김하고 있다. 실내 마스크 착용 의무 해제에 대한 의견의 경우 약 25%의 반대가 존재했다. 이 보고서에서는 이러한 정세에 반하는 위드코로나 정책의 타당성을 확인하고자 한다.

**2) 데이터 분석 설명**

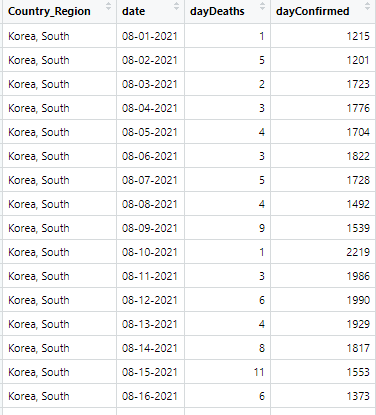
이 보고서에서는 R을 활용하여 한국의 코로나19 확진자 및 사망자 데이터를 분석할 예정이다. 아울러 보다 정확한 분석 및 수치 계산을 위해 기간을 고정하도록 했다. 따라서 ‘2021년 8월 ~ 2022년 7월’로 기간을 고정하여, 해당 1년 동안의 데이터만 활용한다. 데이터를 분석하는 과정에서 파악된 데이터 자체의 기재 오류는 ‘0’ 으로 변환하여 분석했다.

**2. 본론**

**1) 한국의 코로나 환자 추세 분석**

가장 먼저 한국의 일별 코로나 환자 추세 분석을 위한 전처리 과정을 진행했다. 분석 기간 동안 발생한, 한국 코로나 환자의 일별 사망자 수, 일별 확진자 수를 확인한다. 해당 값들은 각각 dayDeath, dayConfirmed이라는 이름의 칼럼으로 지정하였으며, totalkorea라는 데이터 프레임에 속해 있다.

일별 값들을 파악하기 위해, 데이터 파일들을 for문을 통해 모두 불러와 하나의 변수로 합하였다. 동시에 데이터의 일자를 확인할 수 있도록, 새로운 칼럼을 추가하여, 데이터의 일자로 구성되어 있는 파일 명을 기입하였다. 2021년 파일과 2022년 파일을 구분하기 위해 두번의 for문을 만들어 변수를 생성했고 이 후 하나의 변수로 합하는 과정을 거쳤다. 그 결과 아래 결과값과 같이, 한국의 일별 값들을 확인할 수 있다. 상세 코드는 별첨하였다.



데이터를 분석하는 과정에서, 기재 오류가 있는 값들이 존재할 수 있다고 판단했다. 따라서 오기재되어 마이너스(-)값이 나올 경우 해당 값을 '0'으로 변형하는 코드를 추가했다. 또한 오기재된 부분의 위치를 파악하고자 해당 값과 날짜에 해당하는 데이터를 출력하는 코드를 작성했다. 따라서 아래와 같은 코드를 추가해 주었다.

|  |
| --- |
| for(i in 1:nrow(totalkorea)){  if(totalkorea$dayDeaths[i]<0){  totalkorea$dayDeaths[i] <- 0  cat(totalkorea$date[i],"dayDeaths\n")  }  if(totalkorea$dayConfirmed[i]<0){  totalkorea$dayConfirmed[i] <- 0  cat(totalkorea$date[i],"dayConfirmed\n")  }  } |

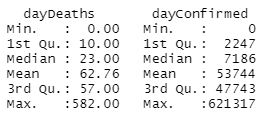
마지막으로 아래 표와 같이, 각 칼럼의 데이터들이 결측치 없이 기재되었는지 확인하고자 하는 코드를 작성하여 값을 확인했다.

|  |  |
| --- | --- |
| # 결측치 확인  summary(totalkorea$dayDeath )  summary(totalkorea$dayConfirmed) |  |

**2) 한국의 코로나 환자 통계**

한국의 일별 코로나 환자의 통계를 위한 분석을 하였다. 코로나 환자에 대한 데이터 분석을 위해 '총 사망자 수', '총 확진자 수'와 '일별 평균 사망자 수', '일별 평균 확진자 수'를 기본으로 도출하였다. 더욱이 해당 값들에 대한 상세 분석을 위해 첨도, 왜도 등의 분석도 추가적으로 진행하였다.

먼저 통계의 기본이 되는 평균, 중앙값, 최대/최소값 파악을 실시했다. 이를 위해 “summary”함수를 사용하였다. ( ※ summary(totalkorea) 코드 이용) 코드에 대한 결과는 아래와 같다.



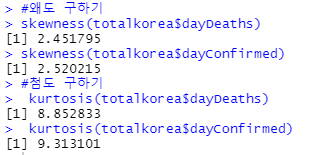
한국의 일별 사망자에 대한 분석 결과 값은 아래와 같다. 조사 기간 동안의 총 사망자는 22908명이며, 일 13명 사망자 발생이 가장 자주 발생했다. 일 최대 사망자 수가 582명인 것과 차이가 존재한다.



한국의 일별 확진자에 대한 분석 결과 값은 아래와 같다. 조사 기간 동안의 총 확진자는 19,616,537명이며, 일 3명 확진자 발생이 가장 자주 발생했다. 최대 발생자 수는 621,317명인 것과 차이가 확연히 느껴지는 숫자이다.



또한 일별 사망자의 왜도는 “2.451795”, 첨도는 “8.852833”이다. 반면 일별 확진자의 왜도는 “2.520215”, 첨도는 “9.313101”이다.

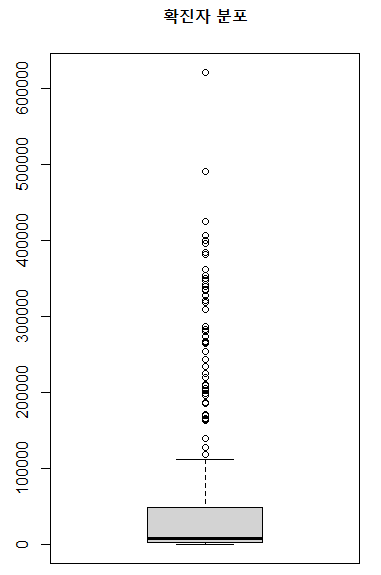
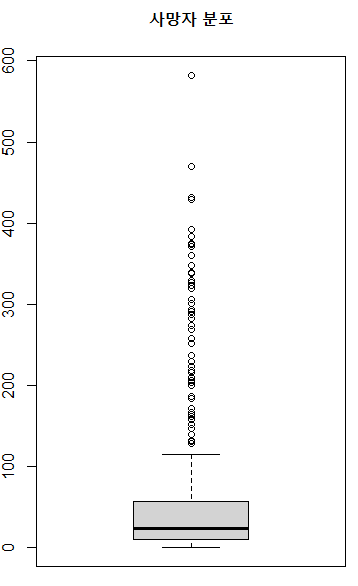


왜도는 분포가 치우친 상태의 정도를 의미한다. 일별 사망자의 왜도는 양수이므로, 왼쪽으로 분포가 치우쳐있다. 즉 평균(62.76) 이상의 값이 다수 존재한다는 것을 알 수 있다. 일별 확진자의 왜도 또한 양수이므로, 평균(53,744) 이하보다 평균 이상의 값이 다수 분포되어 있다. 즉 일별 사망자와 확진자는 최빈값은 낮으나, 평균 이상의 수치가 평균 이하의 수치보다 다양하게 분포되어 있음을 뜻한다.

반면 첨도는 분포가 평균에 모인 정도를 의미한다. 일별 사망자와 일별 확진자의 첨도는 양수이므로, 평균으로 값이 몰려있다. 일별 확진자의 왜도 또한 양수이므로, 평균(53,744) 이상의 값이 다수 분포되어 있다. 다시 말해, 한국의 일별 사망자와 확진자는 평균 이하의 값은 평균과 유사하게 몰려있으며 평균 이상의 값은 다양한 값으로 존재한다는 것을 알 수 있다.

**3) 한국의 월별 코로나 환자 분석**

한국의 월별 코로나 환자 분석에 앞서, 1년동안 대한민국에서 발생한 코로나 발생자수 및 사망자 수 데이터 대상으로 이상치를 파악하고자 한다. 이를 위해 boxplot 함수를 사용하였다.

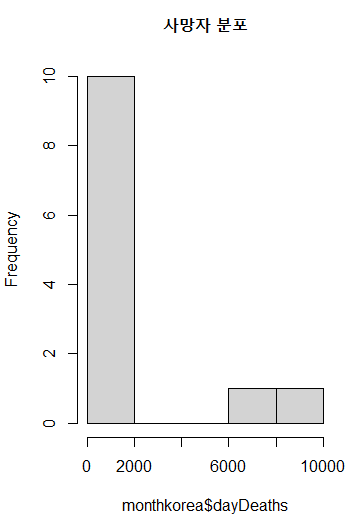
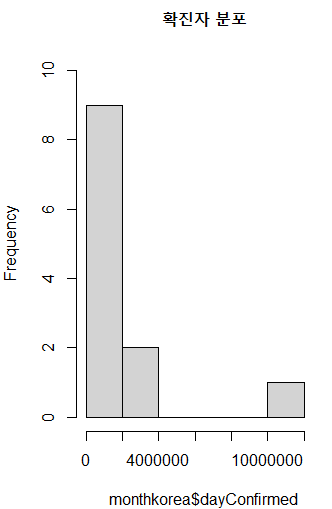
 

한국의 일별 사망자의 IQR은 47, 일별 확진자의 IQR은 45,496이다. (IQR=Q3-Q1) 따라서 일별 사망자의 Maximum은 127.5, 일별 확진자의 Maximum은 115,987이다. 일별 사망자의 Minimum은 -60.5, 일별 확진자의 Minimum은 -43,249이다. 하지만 사망자 및 확진자의 값이 음수가 나올 수 없으므로, Minimum은 0으로 볼 수 있다.

이어서 월별 사망자 및 확진자 분포를 확인하고자 한다. 이는 hist 함수를 사용했으며 코드는 아래와 동일하다.

|  |
| --- |
| hist(monthkorea$dayDeaths,xlim = c(0, 10000), ylim = c(0, 10), main = "사망자 분포")  hist(monthkorea$dayConfirmed,xlim = c(0, 13000000), ylim = c(0, 10), main = "확진자 분포") |

해당 코드로 나타낸 그래프는 아래와 같다.

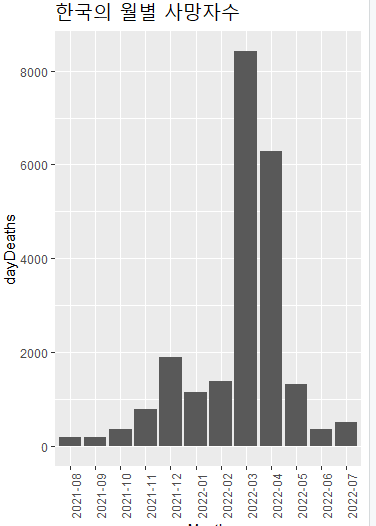
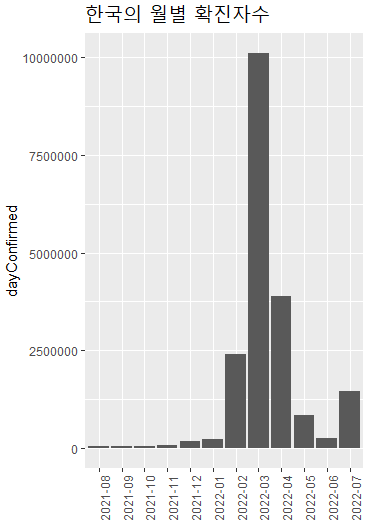
 

한국의 월별 사망자 분포를 확인해 보면, 0~2,000의 분포와 6,000~10,000값을 제외하고는 값이 존재하지 않는 것으로 나타난다. 또한 월별 확진자 분포에서는 4,000,000~10,000,000사이의 값이 존재하지 않는다. 또한 월별 사망자는 0~2,000에서, 월별 확진자는 0~2,000,000에서 값의 산포가 가장 높은 것을 확인할 수 있다. 한국 월별 사망자 분포에서 6,000~8,000과 8,000~10,000의 값이 하나씩 존재한다. 그리고 확진자 분포에서는 10,000,000~12,000,000의 값 하나가 홀로 떨어져 있는 것을 확인할 수 있다. 이는 특이치로써 판단이 가능하다. 하지만 해당 값이 발생한 월을 파악하기에는 어려움이 존재한다.

따라서 보다 직관적인 결과 분석을 위해 ggplot함수를 이용하여 월별 사망자와 확진자 그래프를 그려보았다. 이를 통해 hist 그래프에서 나타난 총 3개의 특이치의 내용을 파악할 수 있다. 코드는 아래와 동일하며, 한국의 월별 사망자 수를 나타낸 그래프는 'grape', 월별 확진자 수를 나타낸 그래프는 'grape2'라는 변수로 코드를 작성했다.

|  |
| --- |
| grape <- ggplot()+ geom\_bar(data=monthkorea, aes(x=Month, y=dayDeaths), stat="identity")  grape <- grape + labs(title='한국의 월별 사망자수')+ theme(plot.title = element\_text(size=18))  grape2 <- ggplot()+ geom\_bar(data=monthkorea, aes(x=Month, y=dayConfirmed), stat="identity")  grape2 <- grape2 + labs(title='한국의 월별 확진자수')+ theme(plot.title = element\_text(size=18)) |

해당 코드로 나타낸 그래프는 아래와 동일하다.

한국의 월별 사망자 수 그래프를 확인해 보면, 2022년 3월과 2022년 4월의 값이 확연히 도드라져 있다. 월별 확진자 수 그래프에서도 유사하게 2022년 3월의 값이 도드라져 나타난다. 이러한 수치를 기록했던 원인은, 당시 한국을 비롯한 세계에 오미크론 변이가 나타나 일시적으로 사망자 수가 증가했던 특이사항이 존재한다.

그 외의 월계에서는 코로나19 사망자 및 확진자의 수치가 상당히 안정적으로 존재하고 있는 것을 확인할 수 있다. 월 사망자 수는 2,000명 이하, 월 확진자 수는 2,500,000명 이하의 수치가 지속적으로 기록되고 있다.

**3. 분석 결과 및 결론**

**1) 데이터 분석 결과 및 결론 도출**

분석 대상 기간 동안의 한국의 코로나19 확진자 및 사망자 수는 본론(3)과 같이 비교적 안정적인 추세에 있다. 2022년 3월 ~ 4월은 급증했던 시기였으나, 이는 오미크론 변이에 의한 수치로 특이 사항으로 볼 수 있다.

데이터 분석 결과, 통계 기간 동안 코로나 확진자 수가 월평균 1,634,711명, 사망자 수가 평균 1,909명 인 것으로 확인된다. 통계 측정 기간의 마지막 월에 해당하는 2022년 7월의 코로나 확진자 수는 1,451,882명, 사망자 수는 513명이다. 즉 월평균 이하의 수치를 기록하고 있다. 코로나19 바이러스는 끊임없이 변이를 하고 있다. 따라서 이를 지속적으로 추적-예방하기에는 다양한 변수가 존재하며, 높은 비용 소모가 예측된다. 더욱이 코로나 확진자 및 사망자 0(Zero)는 불가능에 가깝다고 판단된다. 해외 입출국 자율화로 인한 정성/정량적인 국가 편익에 대한 고려와 실내 마스크 의무 착용, 코로나 지정 병원 등록으로 인한 일반 병실 부족 등으로 인한 불편의성 해소가 필요한 시점이다. 실제로 코로나19 특별 관리 당시, 非코로나19 - 위중증환자의 병실 좌석수 부족 문제가 대두되었다.

이러한 점을 감안하였을 때, 정부의 완화된 코로나19 대응 정책은 긍정적이라고 판단된다. 최대로 발생했던 월의 수치인 10,102,369명에서 마지막 기간인 2022년 7월에는 1,451,882명으로 확진자 수치가 약세화 되고 있는 만큼, 이제는 지난 3년 동안의 국가적 대내외 손실과 개인적 불편의성을 해소할 시기라고 볼 수 있다.

**참고 자료**

1. **데이터셋: Johns’ Hopkis 대학 내 The Center For Systems Science and Engineering(CSSE) :**

<https://github.com/CSSEGISandData/COVID-19/tree/770dafdb73e9dc31140db77b13b1b92cfd8241f9/csse_covid_19_data/csse_covid_19_daily_reports>

1. **[표1] 실내마스크 착용의무 해제에 대한 의견 :**

<http://www.dailypharm.com/Users/News/NewsView.html?ID=296195>