新型コロナウイルスの時系列解析

(3) 移動平均等を用いた感染状況の把握方法について

nino

2023 年 11 月 1 日

１．はじめに

新型コロナの感染状況を把握するため、移動平均などの統計手法が使われている。しかし、移動平均が適切に使用されていないと思われる事例も見受けられた。
そこで、移動平均等の意味を検証するとともに、その結果を参考にして第５波の感染状況を把握する方法について再考察した。

２．統計手法の意味

曜日変動を除去してトレンドを求めるための７日間移動平均や前週同曜日差の意味について検証する。また、７日間移動平均の差分（後述）についても検討する。

2-1　単純移動平均

移動平均（単純移動平均）とは、各時点の移動平均値として、その時点を含む一定期間の平均値を使用する方法であり、次の３つがある[1)]。(英語名はネット検索で調べたもの。)

① 中央移動平均（Central Moving Average、CMA）
② 後方移動平均（Backward Moving Average、BMA）
③ 前方移動平均（Forward Moving Average、FMA）

新型コロナでは主として中央移動平均や後方移動平均が利用されているが、前方移動平均も含めた３つの移動平均の意味と相互の関係について調べた。
７日間移動平均の場合、各日の移動平均値は次のように定義される[1)]。

① CMA：各日とその前後の３日間を含む計７日間の平均値
② BMA：各日とそれ以前の６日間を含む計７日間の平均値
③ FMA：各日とそれ以後の６日間を含む計７日間の平均値

時点 i のデータを x_i とした場合、最新の時点 i からその６日前までの時点［$i-6$］を含む次の計７日間のデータを対象にして、各移動平均を求めた。

x_{i-6},　x_{i-5},　x_{i-4},　x_{i-3},　x_{i-2},　x_{i-1},　x_i

各移動平均は次式で表される。

$$CMA_{i-3} = \frac{1}{7}(x_{i-6} + x_{i-5} + x_{i-4} + x_{i-3} + x_{i-2} + x_{i-1} + x_i) \tag{1}$$

$$BMA_i = \frac{1}{7}(x_{i-6} + x_{i-5} + x_{i-4} + x_{i-3} + x_{i-2} + x_{i-1} + x_i) \tag{2}$$

$$FMA_{i-6} = \frac{1}{7}(x_{i-6} + x_{i-5} + x_{i-4} + x_{i-3} + x_{i-2} + x_{i-1} + x_i) \tag{3}$$

当たり前だが、各移動平均はすべて同じ式（値）となる。しかし、各移動平均値の時点の位置づけが異なり、中央移動平均は時点［$i-3$］に、後方移動平均は時点［i］に、そして前方移動平均は時点［$i-6$］にある。中央移動平均の時点からみると、後方移動平均は３日遅れ、前方移動平均は３日進みとなる。

本来、移動平均の値を当該７日間の中央の時点に位置づけるのが正しい[2、3、4、5)]。それは中央移動平均の時点であり、後方移動平均と前方移動平均は単に時点の位置づけを変えているに過ぎない。単純移動平均の計算は単純で簡単だが、留意すべき点がある。例えば、新型コロナの場合、後方移動平均は最新の時点のデータが常に更新され、移動平均の推移が新規感染者の傾向を反映していると思われがちだが、決してそうではない。また、後方移動平均はデータの変化をより即座に反映し、トレンドやパターンの速い変化を検出するのに適しているわけでもない。７日間移動平均により７日間平均値を求めた場合、その最新の確定値は３日前［$i-3$］にあり、最新までの３日間（時点［$i-2$］から時点 i まで）の値は欠損する[3，5)]。いわゆる、タイムラグは避けられない。
後方移動平均や前方移動平均を用いる場合もあると思われるが、時点の位置づけと解釈に注意する必要がある。

2-2　前週同曜日差

前週同曜日差は、前週の同じ曜日との差であり、曜日変動を除く方法である[1、4)]。
時点 i とその前週同曜日［$i-7$］との差を前週同曜日差 $\triangle WD_i$ とすると、次式で表される。

$$\triangle WD_i = x_i - x_{i-7} \tag{4}$$

2-3　移動平均の差分

感染の動向やピークの時点を把握するためには、移動平均よりも、移動平均の差分（移動平均値による各日とその前日との差であり、１日間の変化量に相当する。）を用いたほうがわかりやすい[4，6)]。
例えば、新規感染者数の移動平均値の差分がプラスなら感染は前日より拡大傾向、マイナスなら減少傾向にあり、また、差分の大小により感染状況が急激にあるいは緩やかに変化しているのかを判断できる。特に、比較的緩やかな変化を示す移動平均のピーク付近の時点は把握しにくいが、移動平均の差分はプラスとマイナスの分岐点すなわちゼロの時点と一致するのでわかりやすくなる。

中央移動平均における時点［$i-3$］とその前日［$i-4$］の差分 $\triangle CMA_{i\text{-}3}$ は、

$$\begin{aligned}\triangle CMA_{i-3} &= CMA_{i-3} - CMA_{i-4} \\ &= \frac{1}{7}\sum_{j=-3}^{3} x_{i+j-3} - \frac{1}{7}\sum_{j=-3}^{3} x_{i+j-4} \\ &= \frac{1}{7}(x_i - x_{i-7})\end{aligned} \tag{5}$$

となり、後方移動平均における時点 i とその前日［$i-1$］の差分$\triangle BMA_i$は、

$$\triangle BMA_i = BMA_i - BMA_{i-1}$$

$$= \frac{1}{7}\sum_{j=-6}^{0} x_{i+j} - \frac{1}{7}\sum_{j=-6}^{0} x_{i+j-1}$$

$$= \frac{1}{7}(x_i - x_{i-7}) \tag{6}$$

となる。同様に、前方移動平均における時点［$i-6$］とその前日［$i-7$］の差分$\triangle FMA_{i-6}$は次式で表わされる。

$$\triangle FMA_{i-6} = \frac{1}{7}(x_i - x_{i-7}) \tag{7}$$

したがって、移動平均の場合と同様に、各移動平均の差分は時点の位置づけが異なるが、すべて同じ値を示す。一方、式(4)より、３つの移動平均の差分は前週同曜日差の 1/7 倍の値に等しい。結局、次式が成り立つ。

$$\triangle CMA_{i-3} = \triangle BMA_i = \triangle FMA_{i-6} = WD_i/7 \tag{8}$$

式(8)より、後方移動平均と前週同曜日差の時点は一致する。各移動平均の差分が前週同曜日差の 1/7 倍の値となったのは、移動平均の差分は１日間の変化量であるのに対し、前週同曜日差はその７倍に相当する７日間の変化量だからである。

３．第５波の感染状況等の把握

これまでの結果を参考に、第５波における感染状況等を把握するための方法について再考察した。内容は前報[7]と一部重複するが、今後の展開に必要なので再掲した。

使用する時系列データは前報と同じ厚生労働省のオープンデータ（毎日の報告値）の次の５項目（略称）である[8]。

①PCR 検査実施人数（検査）、②新規陽性者数（陽性）、③入院治療等を要する者（入院）、④重症者数（重症）、⑤死亡者数（死亡）、　以降は、略称を用いる。

対象期間は、第５波に相当する 2021/7/11(日)～10/23(土)の 105 日間(15 週間)とした。

3-1　報告値と７日間移動平均

５項目について、報告値と各移動平均の関係を調べた。

図１に報告値と各移動平均の時系列を示した。縦軸の単位は人数であり、横軸の日付は１週間間隔である。

まず、報告値と各移動平均の関係について調べた。検査は日毎の検査数が大きく変動しているため、それらの関係はわかりにくい。陽性は、報告値のピークと中央移動平均のピークがほぼ一致し、後方移動平均は中央移動平均より３日遅れ、前方移動平均は３日進みとなっていることが伺える。その傾向は入院でより明瞭となっている。重症も同様な傾向を示すが、死亡は不規則変動が大きいため、やや不明瞭である。

このように、中央移動平均が報告値の経時変化に最もよく対応していた。

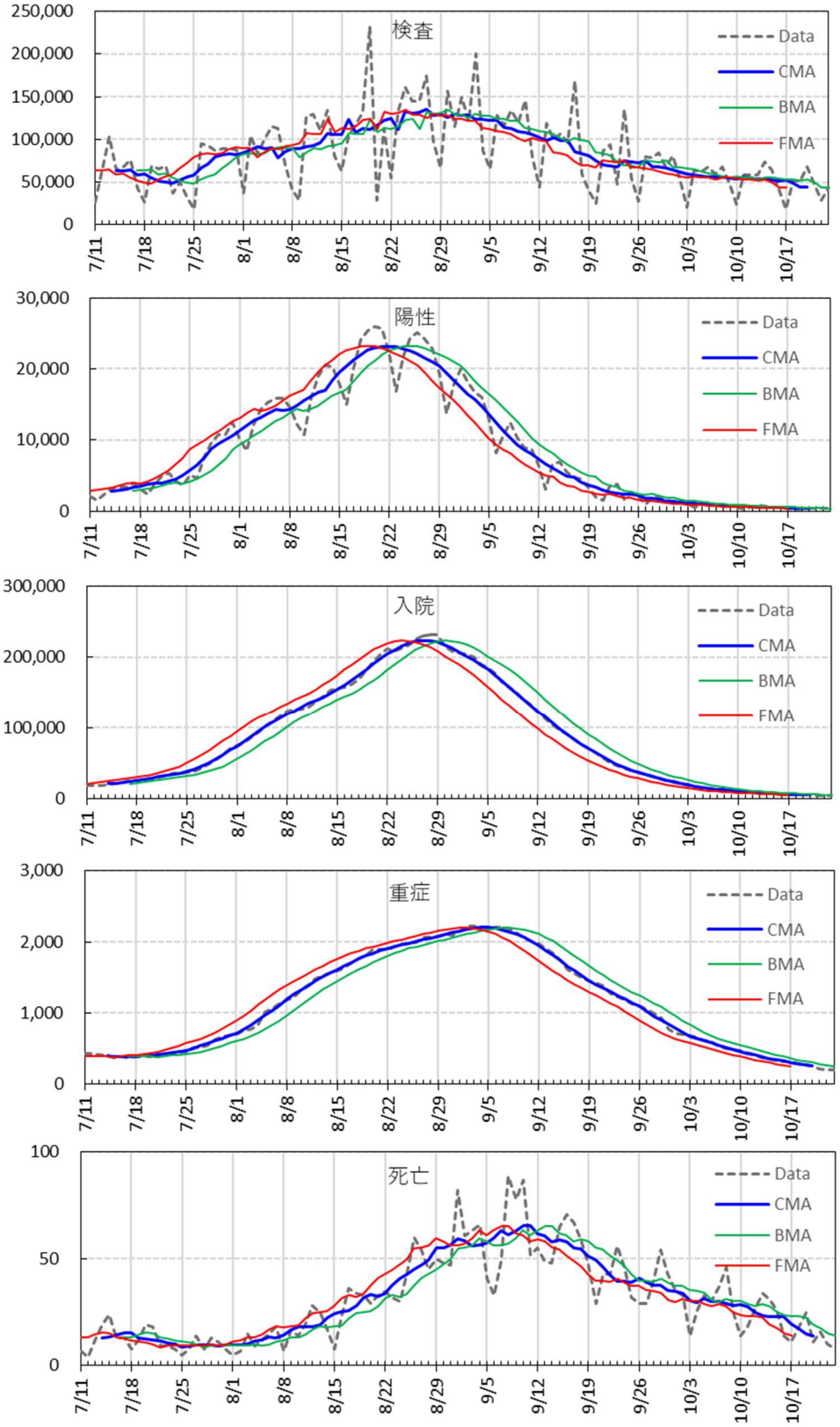

図１　報告値（Data）と各移動平均（CMA、BMA、FMA）の時系列

次に、中央移動平均を用いて経時変化の特徴を調べた。

検査は 8/29 前後にピークが見られた。陽性は当初増加傾向にあるが、8/8 前後に一時停滞気味になった後、再び増加に転じてから、8/22 頃にピークに達した。入院も同様だが、

陽性より滑らかに変化していた。陽性は日毎の感染者数であるため変動が少し大きいが、入院は入院中の人数を積み上げたものであるためと考えられる。また、入院のピークは陽性のピークより数日遅れで現れている。重症は、入院と同様に人数を積み上げたものであるが、一時停滞することはなく増加し、9/5 頃にピークに達した。死亡はピークが 9/5 の週に見られるが、不規則変動が大きいため特定しにくい。死亡も日毎の人数であり、また、人数が他の項目に比べて少ないことが影響していると考えられた。
このように、移動平均だけでは、感染状況等の変化を正確に把握することは難しい。そこで、移動平均の差分と前日同曜日差を用いて検討した。

3-2　移動平均の差分と前週同曜日差について

図２に各移動平均の差分と前週同曜日差の時系列を示した。
まず、各移動平均の差分と前週同曜日差における時点のズレについて確認する。
検査は変動が大きいため時点のズレを把握するのは難しい。陽性については、7/25 の週における３つの移動平均のピークを比較すると、後方移動平均は中央移動平均より３日遅れ、前方移動平均は３日進みであった。この時間のズレは 8/22 前後における３つの移動平均のゼロの時点を比べるとより明瞭に確認できる。一方、前週同曜日差は後方移動平均の経時変化と同期しているが、前週同曜日差の振幅は後方移動平均の振幅の７倍を示した。入院や重症も同様である。このように、先の検証結果が裏付けられた。

次に、中央移動平均の差分を用いて、経時変化の特徴を調べた。
検査は大きく変動しており、特に移動平均のピーク（8/29 頃）前の振幅が大きかった。感染拡大期に、検査数は激しく変動していたことが示唆される。
陽性は、前半はプラスで推移しているが、7/27 から 8/14 頃にかけて落ち込みが見られた。この期間は移動平均で一時停滞している期間と一致していることから（図１）、その状況が差分により顕著に表れた。陽性は 8/14 以降は単調減少し、8/23 になるとマイナスに転じた。感染者数は 8/23 から確実に減少して行ったと考えられる。陽性はその後も減少し続けて、ほぼ 9/5 以降になると減少傾向は緩やかになって行った。したがって、陽性の差分が単調減少し確実にゼロに近づいて行く傾向が認められれば、感染者数は減少に向かっていると推察される。

ただし、この推察は、第５波の全期間の報告値を用いて行われたものである。実際には、その時々の最新までの報告値だけを用いて現況把握や予測を行う必要がある。その最新の移動平均値は、先述したように最新の報告値の時点から３日前にある。言い換えると、感染者数の減少が確実に認識できる時点は、陽性の差分がマイナスを示した 8/23 の３日後の 8/26 となる。このことに留意する必要がある。他項目も同様である。
なお、移動平均の差分がゼロになる時点を事前に予測する方法がある。それは、移動平均の差分が単調減少しゼロに向かっている期間の値を直線近似して外挿することにより、ゼロになる時点を求める方法である。必ずしも精度が良いとは言えず、注意すべき点もあるが、緊急時には有用な情報となるかもしれない。

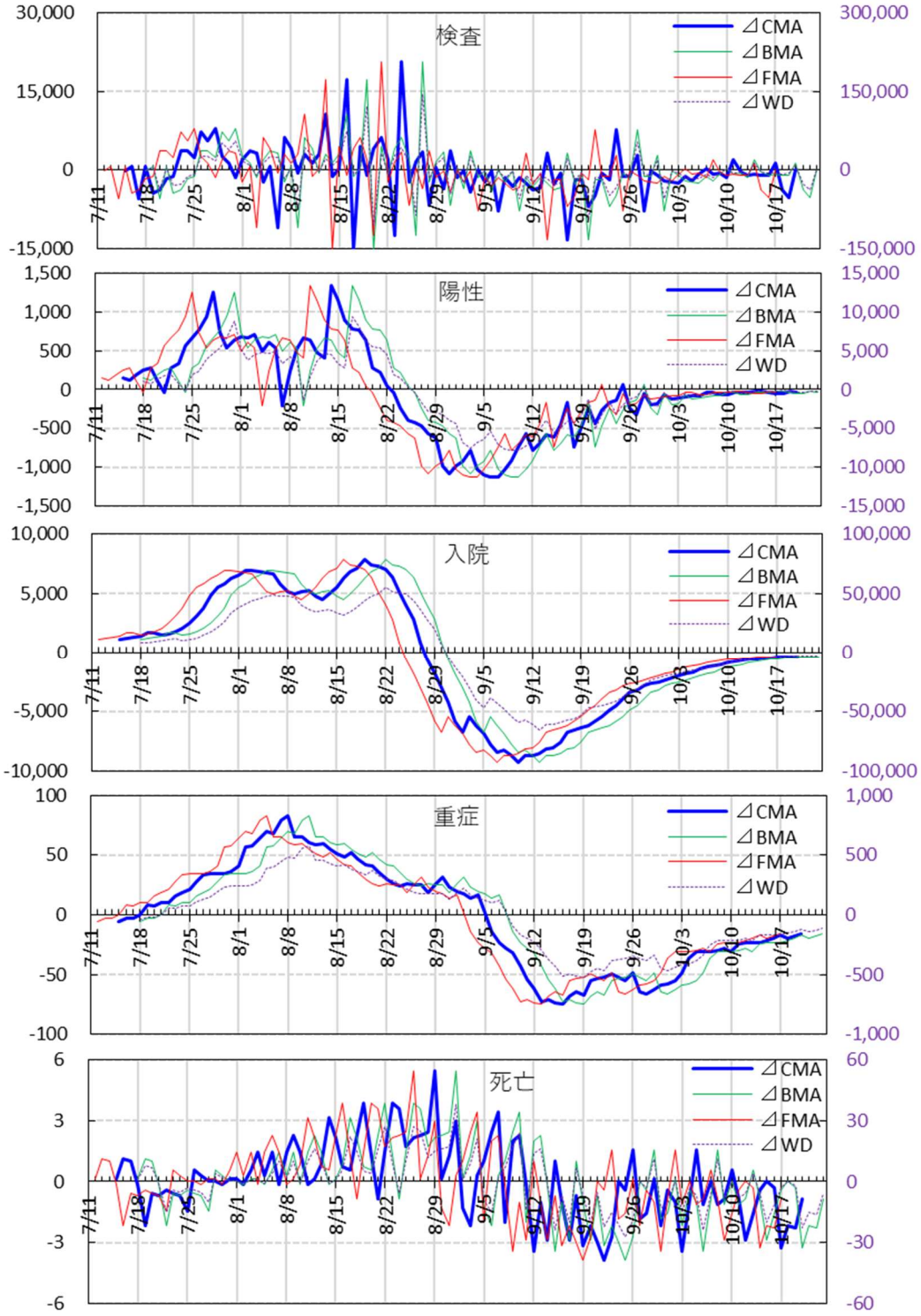

図２　各移動平均値の差分（⊿CMA、⊿BMA、⊿FMA）と
前週同曜日差⊿WD（点線：右目盛）の時系列

入院については、陽性と似た経時変化パターンを示すが、より滑らかでかつ陽性より数日遅れであった。前半途中の落ち込みも見られることから、陽性の影響が及んでいたと考えられる。詳しくみると、入院は 8/20 頃から単調減少し、8/28 になるとマイナスに転じた。したがって、入院者数は 8/23 にマイナスに転じた陽性より 5 日遅れの 8/28 に減少し始め、以降は確実に減少して行った。

重症は、陽性や入院とは異なり前半の落ち込みは無いまま、8/8 以降単調減少して行った。そして、9/5 になるとやや大きく減少した後、9/6 にはマイナスに転じた。このように、入院より 9 日遅れで、重症患者は減少して行った。
死亡は、検査と同様に大きく変動しており、重症より数日遅れの 9/12 前後にマイナスに転じた様子が見られる。死亡の場合は、むしろ移動平均（図１）を用いてピークの時点を判断した方がわかりやすいかもしれない。
これまでに述べてきたような考察ができたのは、第５波における報告値の移動平均の形状が正規分布に類似していたためである。移動平均の形状が異なれば、それに応じて柔軟に解析する必要がある。

５．まとめ

当初、この記事を書く予定はなかったが、追加して書くことにした。というのは、新型コロナで検索すると、「後方移動平均」という言葉が頻繁に使用されていたからである。経済関係などでも使用されているが、その時点の位置づけに疑問を感じたので、その意味を検証した。その内容は先述したとおりである。
ただし、本文の内容が必ずしも十分ではないかもしれない。誤りや説明不足等があれば、ご指摘ください。

参考文献

１）なるほど統計学園
https://www.stat.go.jp/naruhodo/10_tokucho/sonota.html
２）ケンドール統計学用語辞典（1987）　千葉大学統計グループ　丸善
３)物理のかぎしっぽ，周期時系列の統計解析　(3)移動平均とフーリエ変換
http://hooktail.sub.jp/contributions/shuki03.pdf
４）4-4 時系列データ解析　東京大学 数理・情報教育研究センター
http://www.mi.u-tokyo.ac.jp/consortium2/pdf/4-4_literacy_level_note.pdf
５）第４章 季節調整手法
https://www.meti.go.jp/statistics/tyo/iip/riyou-4.pdf
６）福井高専・電子情報工学科　変化の検出・差分処理
https://www.ei.fukui-nct.ac.jp/2018/07/05/diff-processing/
７）新型コロナウイルスの時系列解析　(2) 第５波の特徴
http://hooktail.sub.jp/contributions/covid_02.pdf
８）厚生労働省
https://www.mhlw.go.jp/stf/covid-19/open-data.html