

شناسایی رویدادهای مهم کووید-۱۹ از طریق تویتر

پروژه پایانی کنترل کیفیت آماری-آذر ۱۴۰۰

استاد: دکتر مریم اشرفی

1

مقدمه

“ در طول پاندمی کرونا مردم نشان دادند تمایل دارند زمان بیشتری را در رسانه‌های اجتماعی نسبت به حالت عادی صرف کنند. این میل، پلتفرم‌های رسانه‌های اجتماعی مانند فیس بوک و توییتر را به منبع فعال اطلاعات تبدیل می‌کند که می‌توان با برنامه‌ریزی و اجرای مناسب، اطلاعات ارزشمندی را از این داده‌ها استخراج کرد. در این پروژه قصد داریم با استفاده از نمودارهای کنترل شوارتز، روشی ابتکاری برای شناسایی زمان رویدادهای جدید در دوران پاندمی کرونا ارائه دهیم.



توسعه چشمگیر پلتفرم‌های رسانه‌های اجتماعی

رفتار مردم در شبکه‌های اجتماعی در طول یک بحران

COVID-19

توییتر

2

شرح مسئله

شرح مسئله



توییتز به عنوان منبع برای بررسی افکار عمومی

عوامل موثر بر تعداد توییت ها

روش بررسی

- بررسی هفتگی بجای روزانه

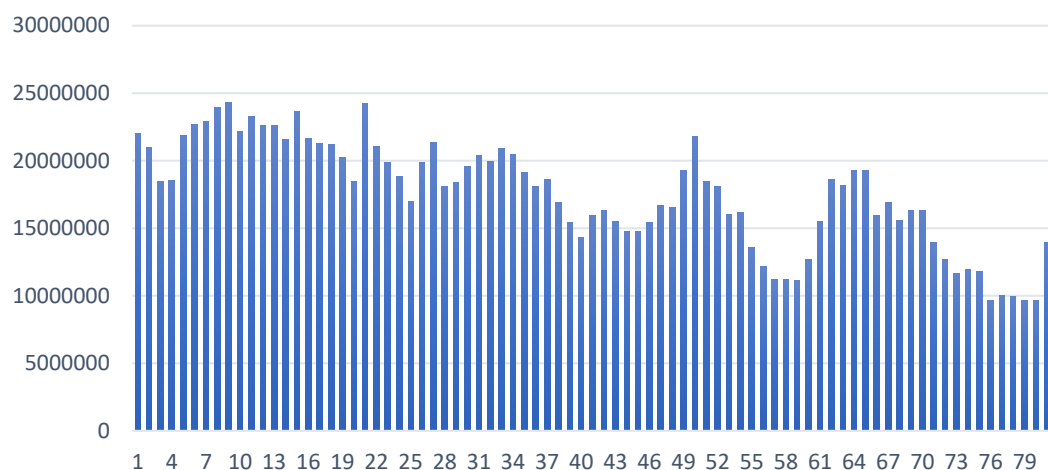
3

روش حل مسئله

روش حل مسئله



تعداد توییت ها



جمع آوری داده

کلیدواژه‌ها

#covid19, #socialdistancing, ...

پیش پردازش و گروه بندی داده

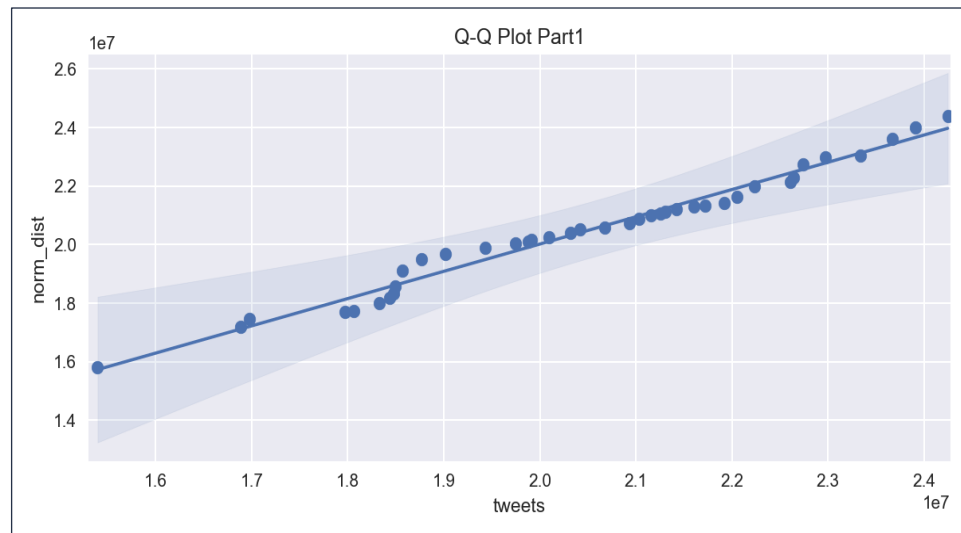
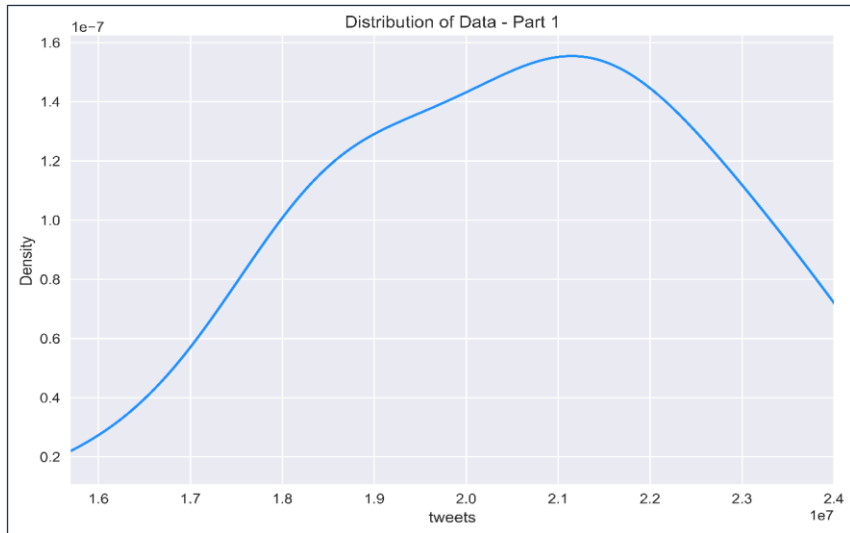
انتخاب نمودار کنترلی

تقسیم دیتاست به دو بخش

4

تحليل قسمت اول دیتاست

بررسی نرمال بودن داده ها



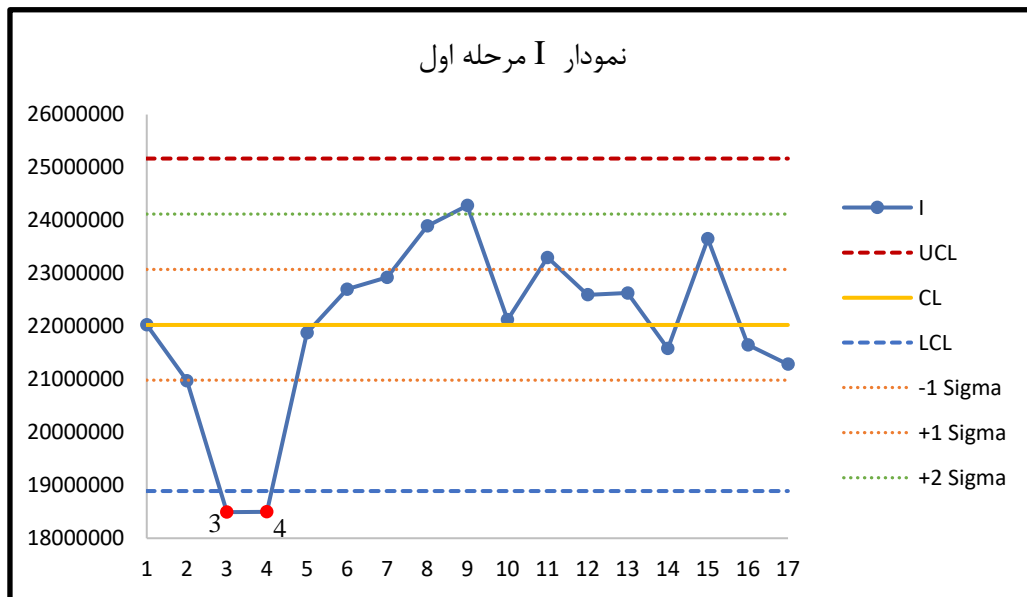
Normality Test Result:

$\alpha = 5\%$

$p - value = 0.67$

\Rightarrow *Dataset is normal.*

فاز اول نمودارها کنترل I/MR

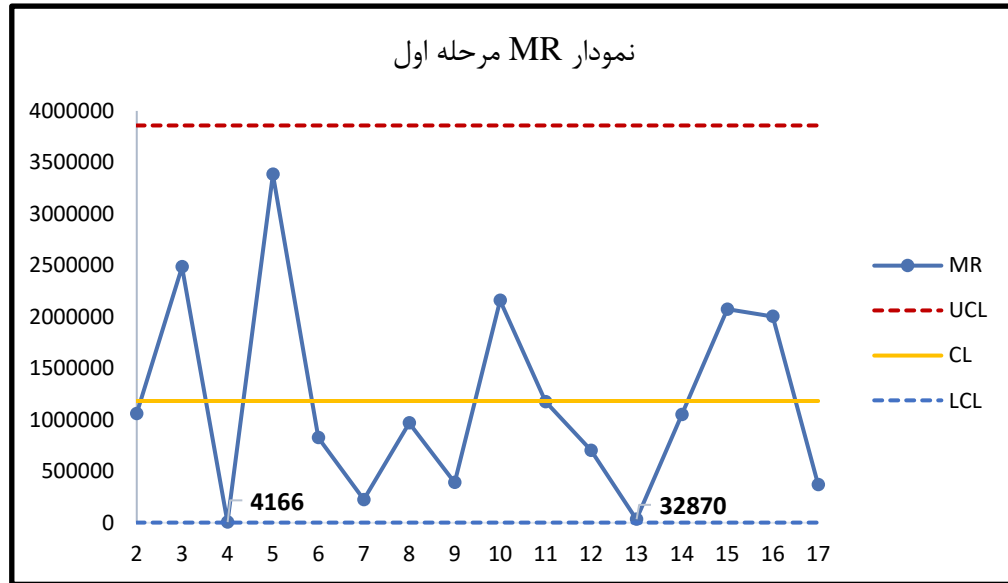


در هفته‌ی سوم و چهارم تقریباً خبر یا حاشیه‌ای داغ درباره ویروس کرونا در خبرگزاری‌ها وجود ندارد و حتی اخباری مبنی بر وجود احتمال بازگشایی کالج‌ها یافت می‌شود.

از هفته‌ی ۵ تا هفته‌ی ۹ شاهد یک‌روند افزایشی هستیم. در بازه‌ی فوق بسیاری از مردم آمریکا در اعتراض به کشته شدن جورج فلویید توسط افسر پلیس شروع به تظاهرات گسترده کردند و موجب افزایش نگرانی‌ها درباره احتمال شیوع دوباره کرونا شدند.

فاز اول نمودارها کنترلی I/MR

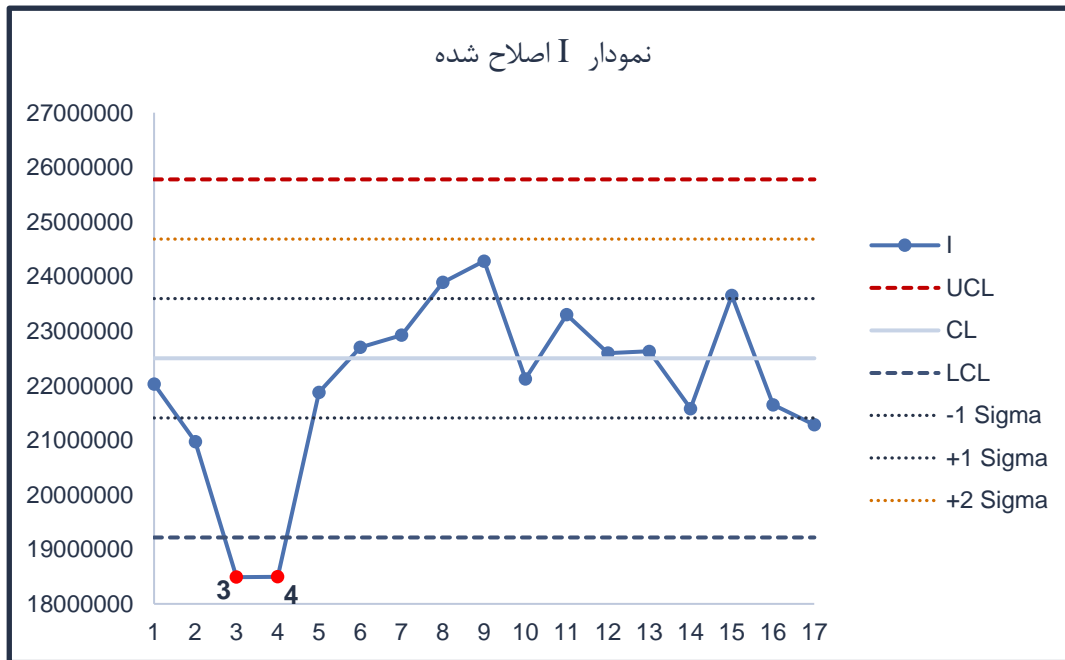
نمودار MR مرحله اول



به دلیل فاصله‌ی زیاد نقاط ۳ و ۴ از حد پایین نمودار، آن‌ها را حذف و به بررسی و محاسبه مجدد پرداختیم.

روند افزایشی هفته ۵ تا ۹ ام با هیچ‌یک از قواعد حساس سازی مغایرتی ندارد، در نتیجه اقدامی انجام نمی دهیم.

فاز اول نمودارها کنترلی I/MR

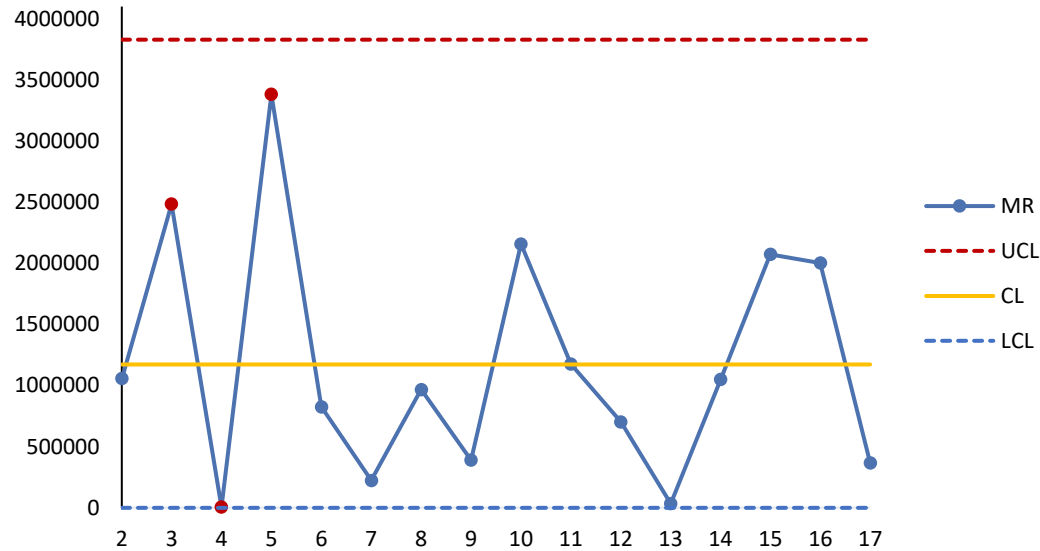


همان‌طور که ملاحظه می‌شود پس از حذف دیتاهای ۳ و ۴ از محاسبات سایر نقاط در میان حدود قرار گرفته است و این به معنای پایان فاز اول محاسبات و آغاز فاز دوم می‌باشد.

حدود کنترل چارت I		
LCL	CL	UCL
۱۹۳۸۳۴۵۴	۲۲۵۰۰۴۸۸	۲۵۶۱۷۵۲۲

فاز اول نمودارها کنترلی I/MR

نمودار MR اصلاح شده

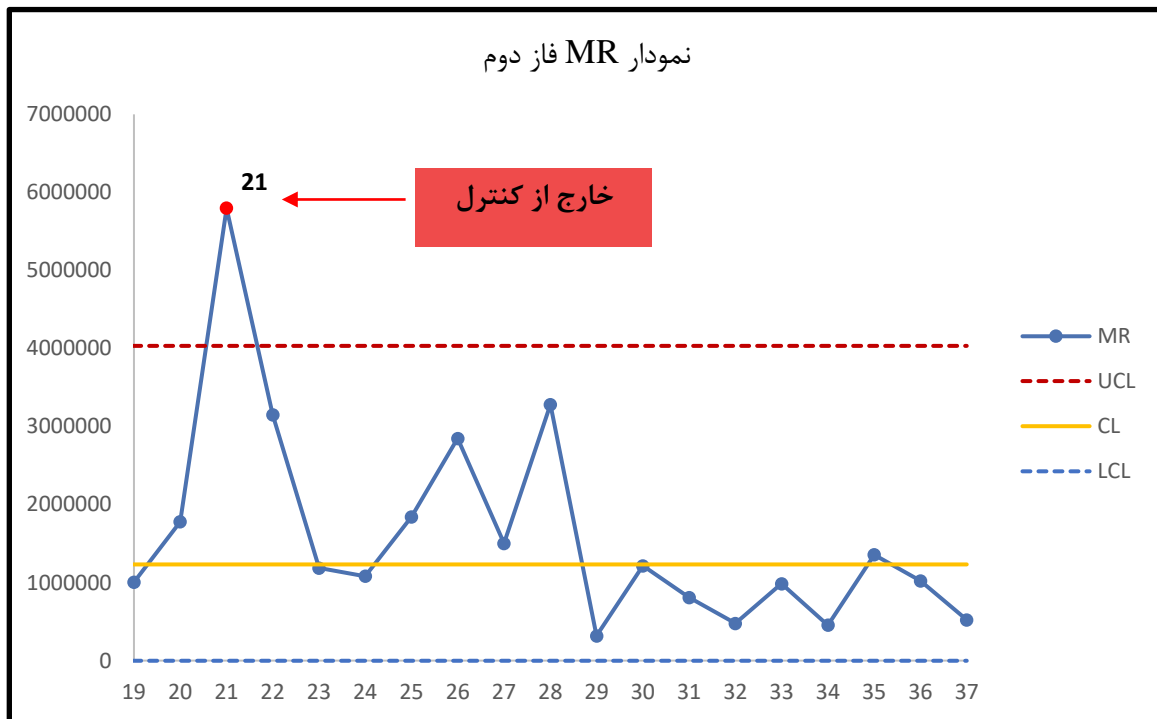


حدود کنترل چارت MR		
LCL	CL	UCL
۰	۱۱۷۲۰۰۵	۳۸۲۸۹۳۹

فاز اول نمودارها کنترل I/MR

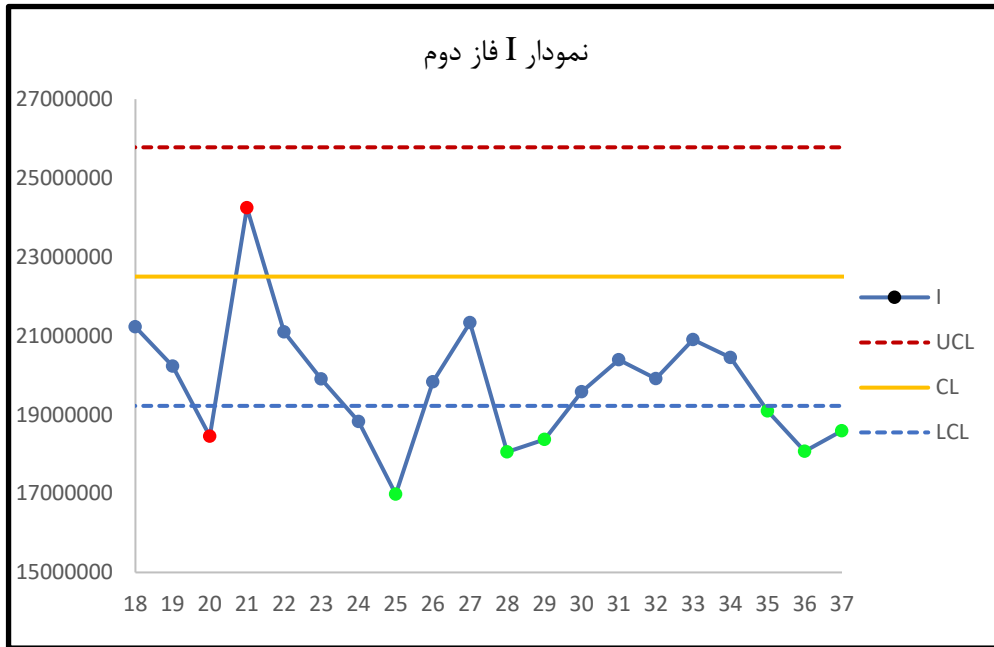


نمودار MR فاز دوم



هشدار دریافتی، مربوط به اوایل اکتبر ۲۰۲۰ است. پس از جست‌وجو در اخبار مشخص شد، در بازه‌ی فوق تست کرونای دونالد ترامپ رئیس‌جمهور وقت آمریکا در بازه‌ی حدوداً یک ماه مانده به انتخابات ۲۰۲۰ آمریکا، مثبت شده و باعث ایجاد موجی از توییت‌ها در رابطه با این مضمون شده است.

فاز دوم نمودارها کنترلی I/MR



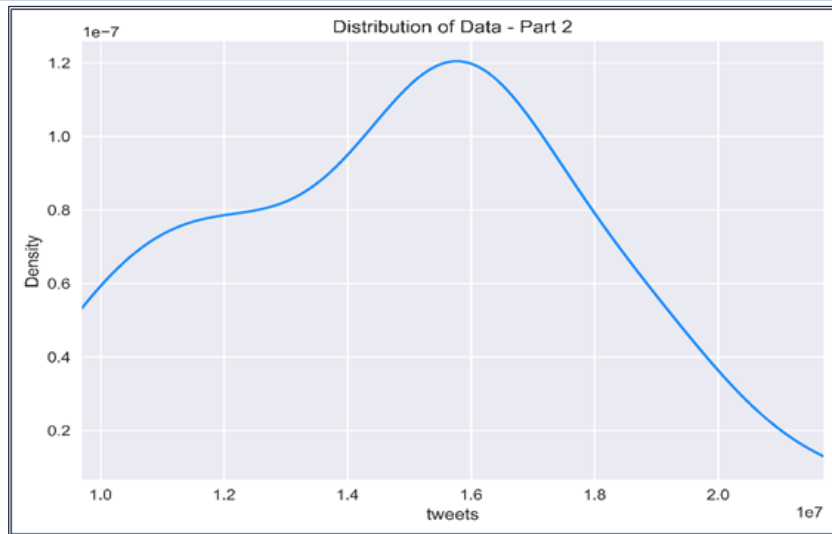
اهمیت مردم به مقوله‌ی کرونا به شدت کاهش یافته، به طوری حتی درصد رعایت پروتکل‌های بهداشتی توسط مردم کاهش چشم‌گیری داشته است.

اثر اتفاقات خاص و حاشیه ای

5

تحليل قسمت دوم دیتاست

بررسی نرمال بودن داده ها



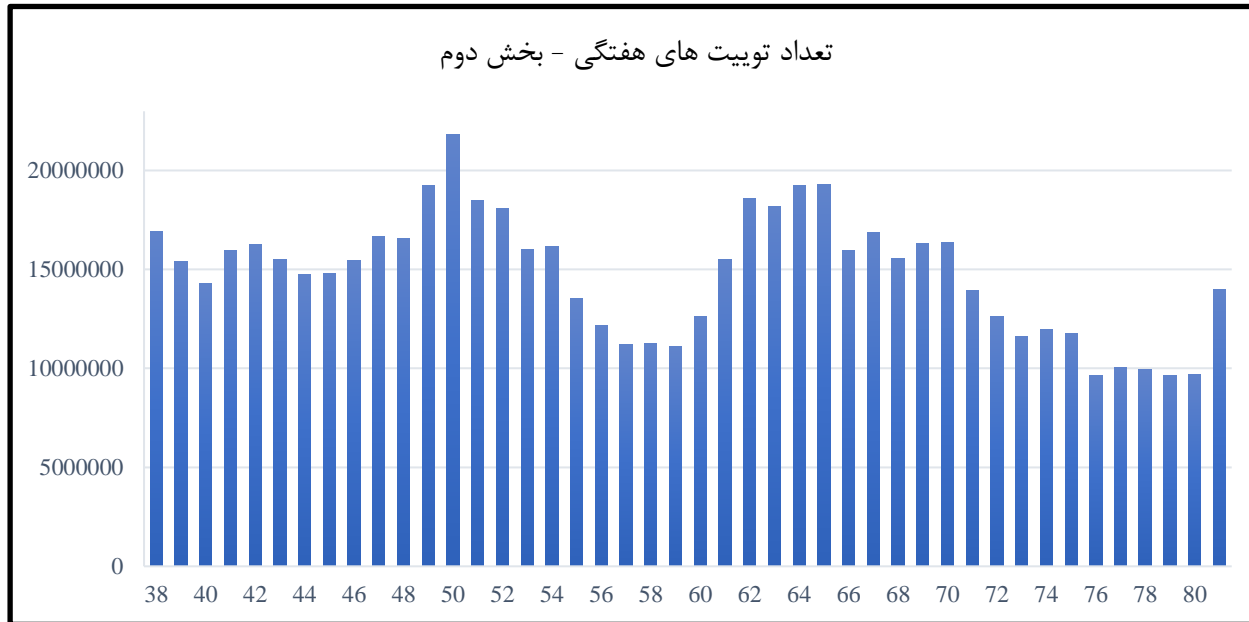
Normality Test Result:

$\alpha = 5\%$

$p - \text{value} = 0.56$

\Rightarrow *Dataset is normal.*

نمودار میله‌ای بخش دیتاست

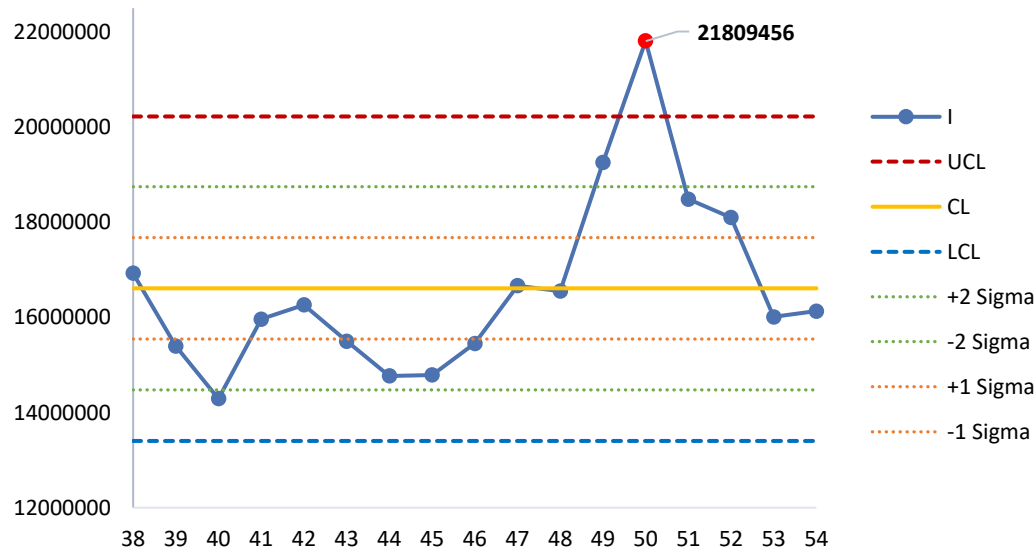


مشاهده یک روند سینوسی طی فرآیند که
حاکی از تحت کنترل نبودن فرآیند است.

فاز اول نمودارها کنترلی I/MR



نمودار کنترلی I فاز اول



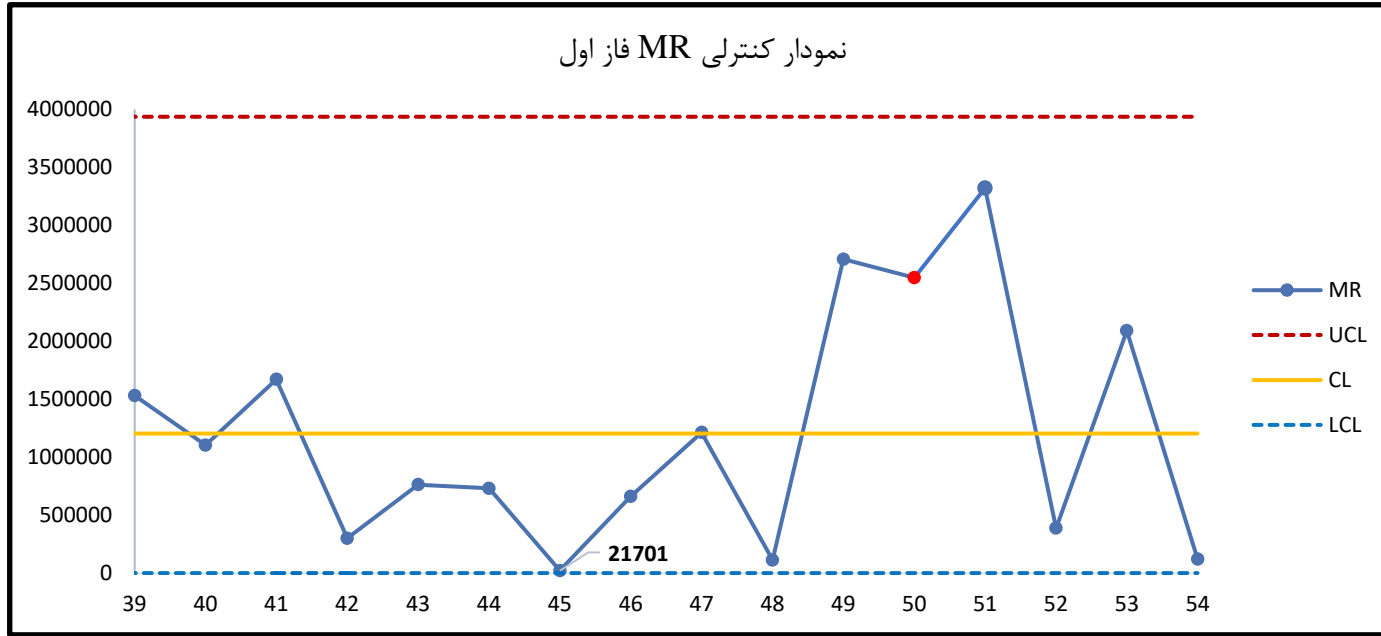
از هفته ۳۹ تا ۴۶: ۸ نقطه متوالی پایین CL

هفته ۵۰: تعداد جانباختگان و مبتلایان طبق آمار WHO

بیشینه شده است. این روند صعودی هفته ۴۹ را نیز شامل

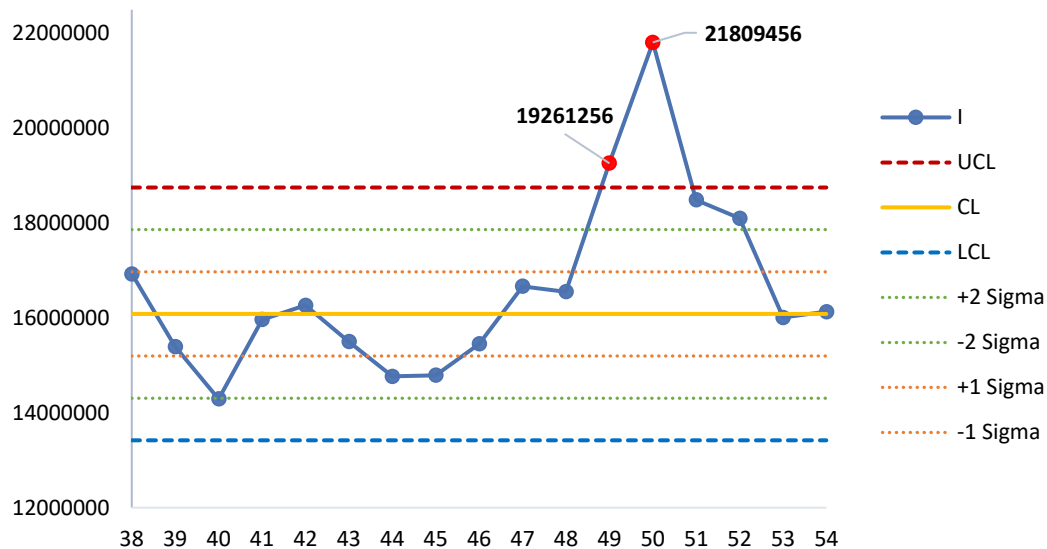
میشود. هر دو نمونه از محاسبات حذف میشوند.

فاز اول نمودارها کنترلی I/MR



فاز اول نمودارها کترلی I/MR

نمودار کترلی I اصلاح شده فاز اول

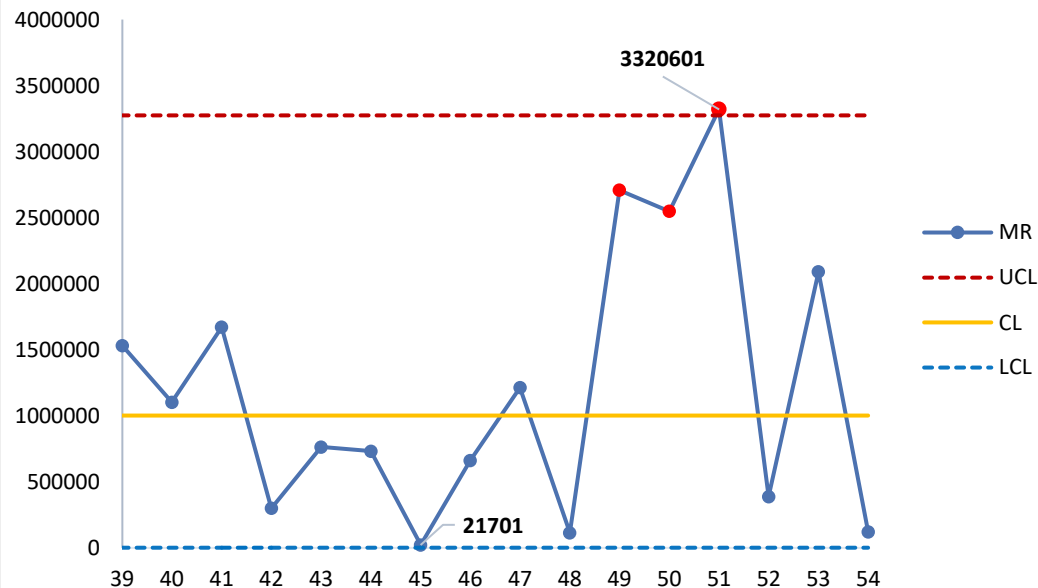


حدود کترل چارت I

LCL	CL	UCL
۱۳۴۲۲۴۹۱	۱۶۰۸۷۹۷۹	۱۸۷۵۳۴۶۷

فاز اول نمودارها کنترلی I/MR

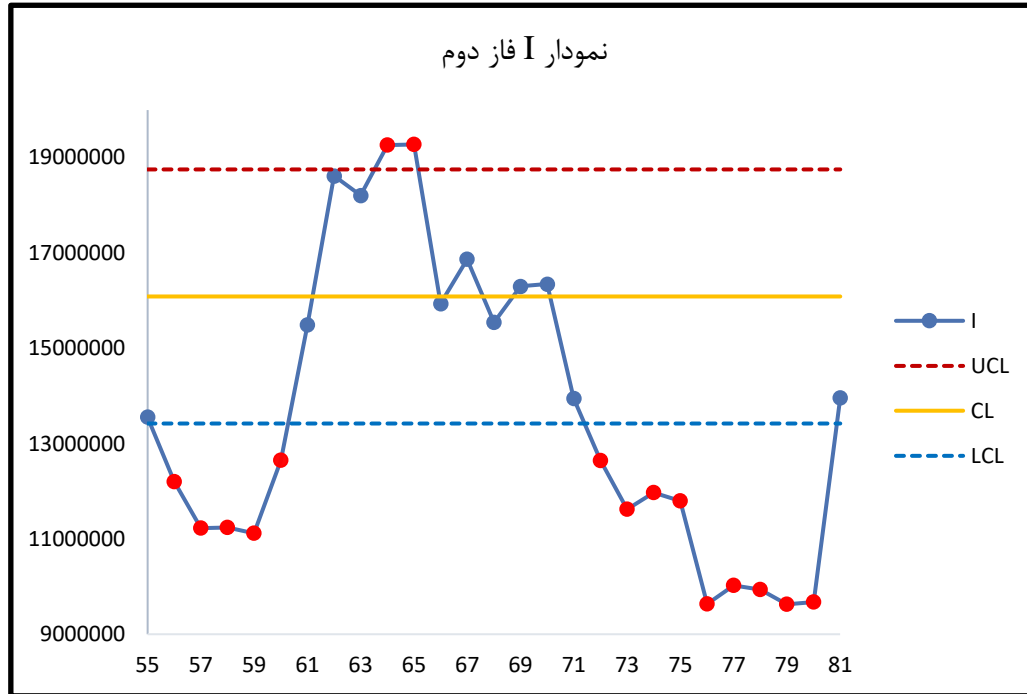
نمودار کنترلی MR اصلاح شده فاز اول



با رسم حدود تصحیح شده می بینیم که نمونه ۵۱ از حد UCL نمودار MR رد شده، اما با بررسی بیشتر دلیلی برای این اتفاق پیدا نمی کنیم، از طرف دیگر به دلیل اینکه فاصله از نمونه تا حد UCL زیاد نیست، از این تفاوت چشم پوشی کرده و حدود را اصلاح نمی کنیم و به فاز دوم می رویم.

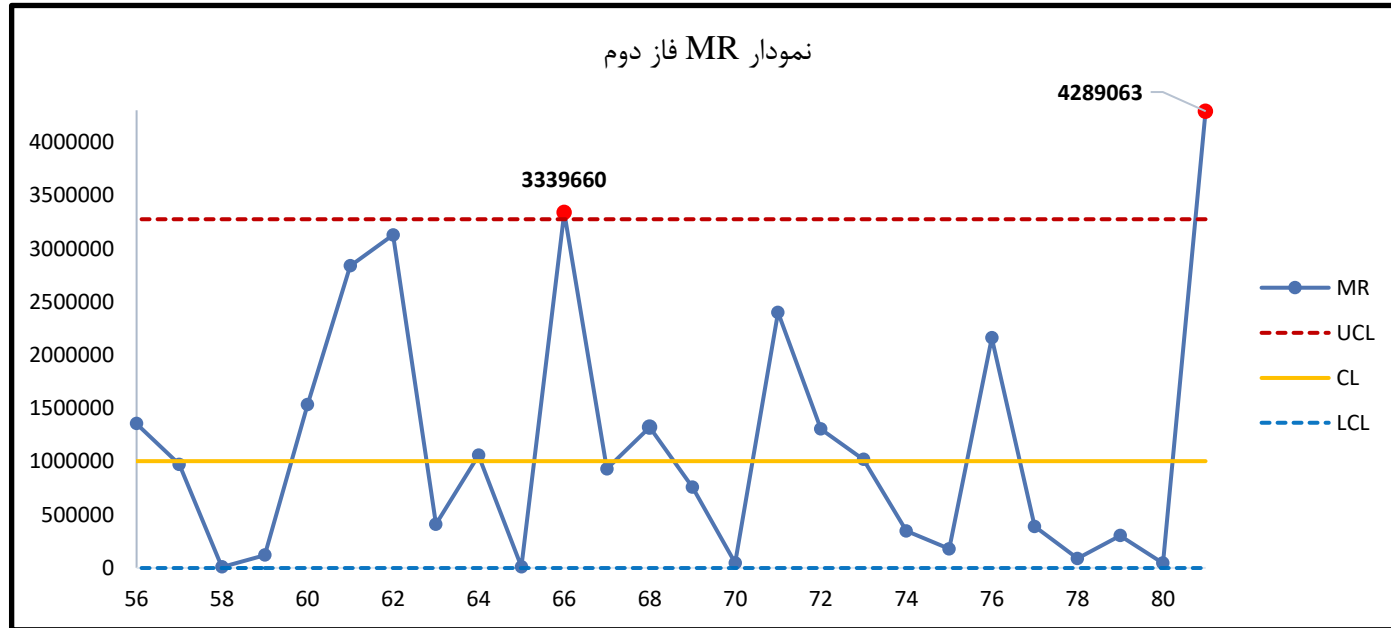
حدود کنترل چارت MR		
LCL	CL	UCL
۰	۱۰۰۲۲۲۳	۳۲۷۴۲۶۴

فاز دوم نمودارها کنترلی I/MR



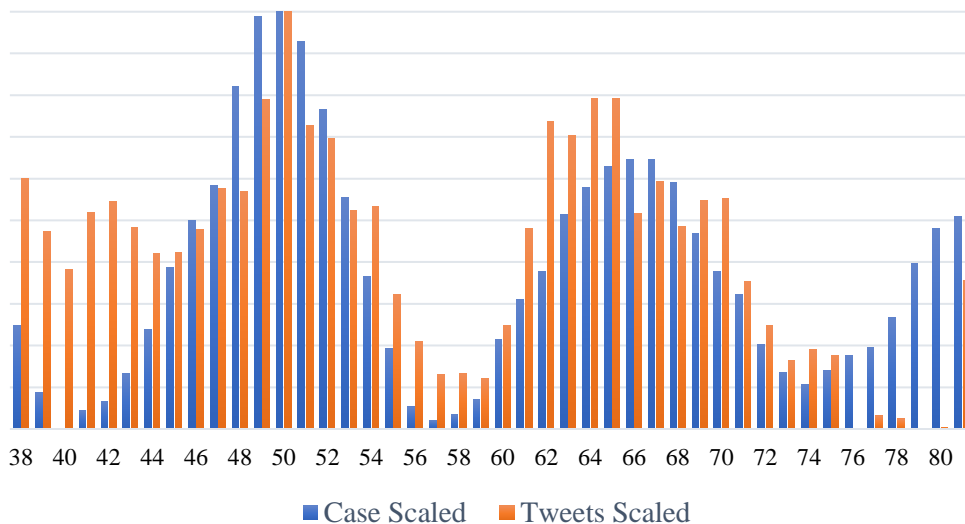
همان گونه که انتظار داشتیم، به علت روند سینوسی که در بخش دوم وجود دارد، نمونه‌های زیادی خارج از حدود افتاده‌اند. در گام بعد به پیدا کردن علت سینوسی بودن تعداد توییت‌ها از هفته ۳۸ به بعد می‌پردازیم. به این منظور، روند کلی تعداد توییت‌ها را با ابتدا با شمار مبتلایان و سپس با شمار فوتی‌ها که در داشبورد سازمان بهداشت جهانی قرار گرفته است، مقایسه می‌کنیم.

فاز دوم نمودارها کنترلی I/MR



نمودار میله ای مقایسه روند مبتلایان با تعداد توئیت ها

روند توئیت ها در برابر مبتلایان



برای مقایسه بهتر تعداد مبتلایان و تعداد توئیت ها در نمودار میله ای، با استفاده از روش Min-Max Normalization، داده ها را اسکیل می کنیم .

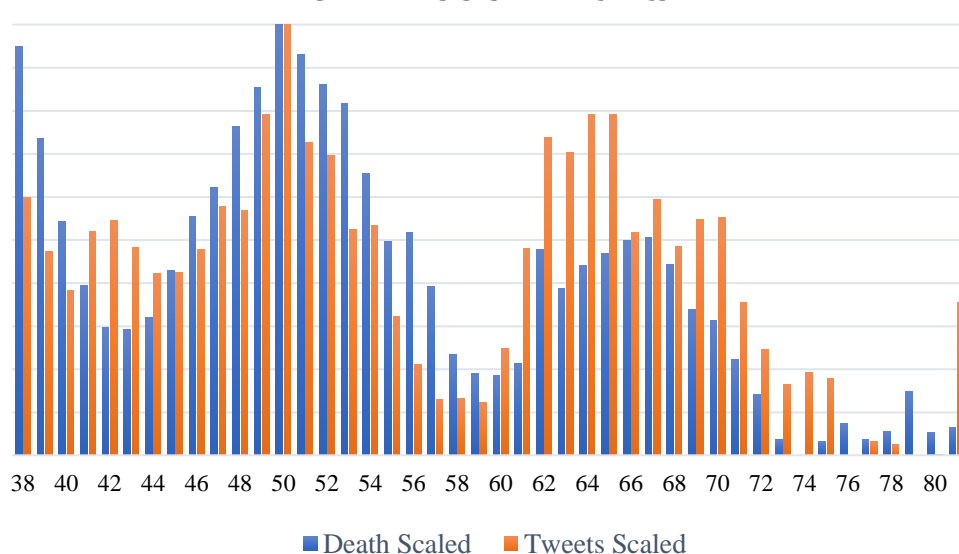
همان طور که از نمودار مشهود است، در بسیاری موارد با افزایش تعداد مبتلایان کرونا، شمار توئیت ها نیز افزایش میابد اما هفته هایی نیز وجود دارد که این موضوع را رد می کند.

ضریب همبستگی بین این دو برابر ۶۵.۹۱٪ است که همبستگی خوبی را نشان می دهد.

نمودار میله‌ای مقایسه روند جانباختگان با تعداد توئیت‌ها



روند توئیت‌ها در برابر جانباختگان



برای مقایسه بهتر تعداد مبتلایان و تعداد توئیت‌ها در نمودار میله‌ای، با استفاده از روش Min-Max Normalization، داده‌ها را اسکیل می‌کنیم.

از این نمودار نتیجه می‌گیریم که روندهای سینوسی که در هر یک از نمودارها وجود دارد، کاملاً بر همدیگر منطبق هستند و با کاهش تعداد توئیت در هفته، تعداد جانباختگان کاهش و با افزایش تعداد توئیت، تعداد جانباختگان افزایش می‌یابد.

ضریب همبستگی بین این دو برابر ۷۴.۷۹٪ است که نشان می‌دهد تعداد توئیت و جانباختگان را به شدت همبسته است.

“

هدف از مطالعه فوق بررسی ارتباط بین تعداد توییت های مرتبط با کرونا و وقایع اتفاق افتاده در دوران همه گیری کرونا بود. از بخش های قبل می توان نتیجه گرفت، تعداد توییت ها و اتفاقات، ارتباط نزدیکی بایکدیگر داشته و می توان از روی تغییرات یکی از آنها، تغییرات دیگری را تا حدودی پیش بینی کرد.



با تشکر

تهیه کنندگان:

پدرام پیرو اصفیا

مهدیار صادقی

سپیده کریمی

مهدی محمدی