



**تحلیل آماری داده‌های ویروس جدید کرونا
(کووید ۱۹)**

محمد اسماعیل حنی

۲۹ / خرداد ۱۳۹۸

مدلسازی انتشار کرونویروس در ایران

مقدمه

در این نوشتار قصد داریم با ارائه یک مدلسازی، رفتار ویروس کووید ۱۹ (ویروس جدید کرونا) را در داخل مرزهای سیاسی ایران تحلیل و بررسی کنیم. این مدلسازی از جنس دنباله‌های مارکوف است. یعنی یک رشته اپیزود می‌سازیم که هر قسمت از دنباله، تنها به قسمت قبلی آن وابسته است. در هر مرحله بیماران و کشته‌ها مورد مطالعه قرار می‌گیرد و در نهایت نیز یک نتیجه‌گیری از روند رشد این بیماری خواهیم داشت.

بخش ۱، فرض‌های ساده‌سازی

در این مدلسازی مفروضات ساده‌سازی زیر انجام شده است:

1. ایران ۸۳ میلیون جمعیت دارد
2. ایران با تراکم یکنواخت و به شکل مربع می‌باشد.
3. هر فرد در ارتباطات روزانه‌اش به استفاده از تابع نرمال ۲ بعدی، به صورت تصادفی افرادی نزدیک به خودش را انتخاب می‌کند و با آن‌ها مرادده می‌کند.

پارامترهای اپیدمی

هر مدلسازی نیازمند یک سری مقادیر ثابت هست که در این مدل نیز، پارامترهایی مؤثر در قوت و ضعف اپیدمی را باید لحاظ کنیم. مقادیر ثابتی که در مدلسازی استفاده شده است در ذیل به تفصیل بیان می‌شوند.

1. احتمال ابتلای کسی که یکبار با یک بیماری کرونایی در ارتباط است حدود ۰/۰۳۷ است. (این عدد با مشورت دوستان درگیر با کادر درمانی و با توجه به ۲۰ روز نخست بیماری در ایران انتخاب شده است.)
2. احتمال مرگ بر اثر کرونا ۲٪ و پس از تغییر سیاست در تعیین فوتی‌های بیماری‌های تنفسی به عنوان کرونا ۱٪، به منظور لحاظ کردن متبحر شدن و شوک اولیه جامعه پزشکی احتمال مرگ + $\frac{1}{r}$ می‌شود.
3. **(*تصحیح شده)** دوره مخفی بیماری در ایران: ابتدا با فرض وجود ۱۰ روز مخفی از بیماری تحلیل‌ها را شروع کردیم ولی بعداً متوجه شدیم که عدم در نظر گرفتن دوره مخفی باعث انطباق بیشتر نمودار با داده‌های رسمی می‌شود. لذا این دوره را نادیده گرفتیم. احتمالاً دلیل این اتفاق این است که نمودار ما ۱۰ روز جلوتر از داده‌های واقعی است (در این تحلیل ناقلین تخمین زده می‌شود نه افراد بستری).
4. **(*تصحیح شده)** بیماری با ورود حدود ۴ فرد آلوده در ایران منتشر شده است. (دلیل: با توجه به قرنطینه شدن اکثر ورودی‌ها از چین، نباید تعدادی زیادی آلوده وارد کشور شده باشد. با توجه به آلودگی مستقل گیلان و قم،

حداقل دو فرد آلوده وارد کشور شده بودند. با توجه به اینکه بیمار شماره صفر ایران شناسایی نشده است لذا فرض ۴ آلوده نباید فرض سنگینی باشد).

5. طول دوره بیماری حداکثر ۴۵ روز است. (دلیل: پرسش از پزشک)

6. بیمار از روز ۱۵ پس از آلودگی ممکن است بمیرد. (دلیل: ۵ روز نخست صرف کمون و تشدید بیماری می‌شود).

7. بیمار از روز ۱۸ پس از آلودگی ممکن است درمان شود. (دلیل: متوسط ۵ روز کمون + ۳ روز نقاهت) [دقت

کنید در این مدلسازی منظور از بیماری، آلوده شدن به ویروس است و ناقلین نیز بیمار محسوب می‌شوند، همچنین منظور از درمان ناتوانی در انتقال است لذا در دوره نقاهت فرد هنوز بیمار محسوب می‌شود]

8. **(*تصحیح شده)** متوسط ارتباطات مردمی: (دلیل: اعداد این بخش به صورت حدسی و با شناخت اجتماعی تنظیم شده است).

- **(جدید)** افراد در قرنطینه: حداکثر با یک نفر ارتباط غیرایمن دارند آن‌هم با احتمال کمتر از یک دهم
- ۱۰ روز آخر بهمن و ۱ اسفند: به طور متوسط ۲۵ نفر (هر چند جامعه شامل فروشندگان و معلمین هست که با جمعیت زیادی درگیر هستند ولی زنان خانه‌دار نزدیک به ۲۵٪ جامعه را تشکیل می‌دهند. کودکانی که نزد مادران زندگی می‌کنند، برنامه‌نویس‌ها، افراد کم‌توان حرکتی و... ارتباطهای کمی دارند).
- ۱۰ روز اول اسفند: به طور متوسط ۱۲. با توجه به انتخابات ۲ اسفند و تعطیلی اکثر مدارس کشور از ۳ اسفند (به مرور) می‌توان این عدد را عدد مطلوبی در نظر گرفت
- ۱۰ روز دوم اسفند: متوسط ۸ نفر. حساسیت عمومی افزایش پیدا کرده. برخی شرکتهای خصوصی نیز به دورکاری مشغول شدند و همه ارگان‌های آموزشی تعطیل هستند.
- ۱۰ روز سوم اسفند: به طور متوسط ۱۰ نفر. به دلیل خرید شب عید این عدد رشد می‌کند.
- روز چهارشنبه سوری (روزهای ۳۰ام و ۳۱ام بیماری): به دلیل وجود افراد بی‌مسئولیت این روزها تا ۱۲ نفر ارتباط افزایش پیدا می‌کند.
- ۲ روز اول فروردین: به طور متوسط ۱۵ نفر. به دلیل دید و بازدیدها
- تا ۷ فروردین: ۶ نفر
- ادامه: در صورت اعمال قرنطینه، ۳ نفر. دقت کنید متوسط جمعیت خانوارهای ایرانی بین ۳ تا ۴ نفر است. در صورتی که تنها یک نفر مامور خرید مایحتاج روزانه شود عدد ۳ عددی منطقی است.

9. **قرنطینه افراد:** فرض بر این است در روزهای نخست تنها بیماران رسمی قرنطینه می‌شوند ولی به مرور زمان و افزایش تکنولوژی و تبحر پزشکی ناقلین نیز قرنطینه خواهند شد. فرض بر این است که جمهوری اسلامی ایران ظرف یکسال به صورت خطی به تکنولوژی قرنطینه کردن ۱۰۰٪ ناقلین دست پیدا خواهد کرد.

10. **ضریب داده‌ها:** داده‌های مربوط به بیماری کووید ۱۹ به سه دسته تقسیم می‌شوند

- داده‌های رسمی. آنچه که هر روز ساعت ۱۳، وزارت بهداشت جمهوری اسلامی ایران بیان می‌کند.
- داده‌های سازمانی، در اختیار نهادهای خاص است و مردم اطلاعی ندارند. ممکن است داده‌های رسمی منطبق بر داده‌های سازمانی باشد، اما به هر حال داده‌های سازمانی حاوی جزئیات بیشتری است.
- داده‌های واقعی، هیچ فردی از این داده‌ها اطلاعی ندارد. افراد مبتلای بدون نشانه، مبتلایانی که به پزشک مراجعه نمی‌کنند یا به دلیل اشتباه در تشخیص اولیه پزشک تست کرونا روی آن‌ها انجام نمی‌شود. با تحلیل طولانی که در پیوست این مقاله به آن اشاره می‌شود، متوجه می‌شویم که تعداد بیماران واقعی حدود ۸/۲ برابر داده‌های رسمی است. در این مقاله از همین تخمین استفاده خواهیم کرد و از داده‌های تخمینی استفاده خواهیم کرد.

بارگذاری:

بارگذاری: همه ۸۲ میلیون ایرانی را روی صفحه 1×1 می‌چینیم. از بین همه ایرانیان ۱۰ نفر به تصادف انتخاب و آن‌ها را آلوده می‌کنیم.

قسمت k ام (اپیزود): هر بیمار با توجه به k با n نفر دیدار می‌کند. با احتمال 0.18 هر کدام را آلوده می‌کند.

نتیج:

در دو نمودار ۱ و ۲ (صفحه بعد)، نمودار داده‌های تخمینی (داده‌های واقعی * $8/2$) را به رنگ آبی و داده‌های مدل‌سازی را با خط چین نارنجی مشاهده می‌کنید. سطر پایینی هر کادر شامل پیشبینی هست که مدل‌سازی ما ارائه می‌کند.

با توجه به زمانبر بودن مدل‌سازی سناریوی قرنطینه تا ۱۶۲ روز شبیه‌سازی شده است که در تصویر مشاهده می‌کنید:

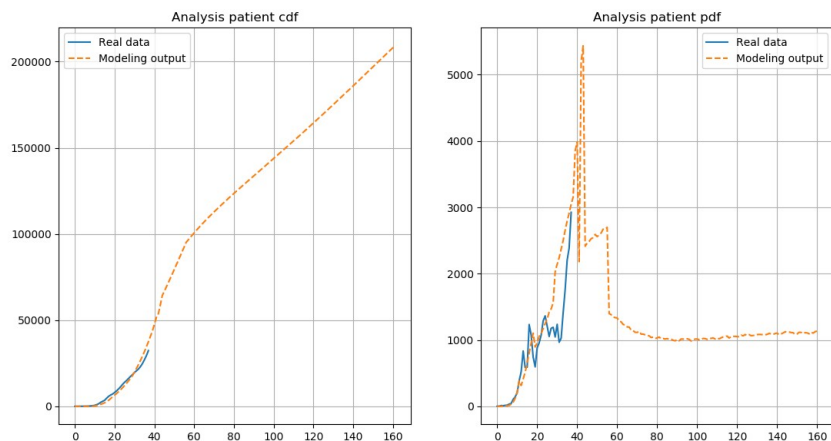
مجموع افراد ناقل: ۱۷۱۵۴۰۶ نفر

مجموع افراد فوتی: ۴۱۲۷۵ نفر

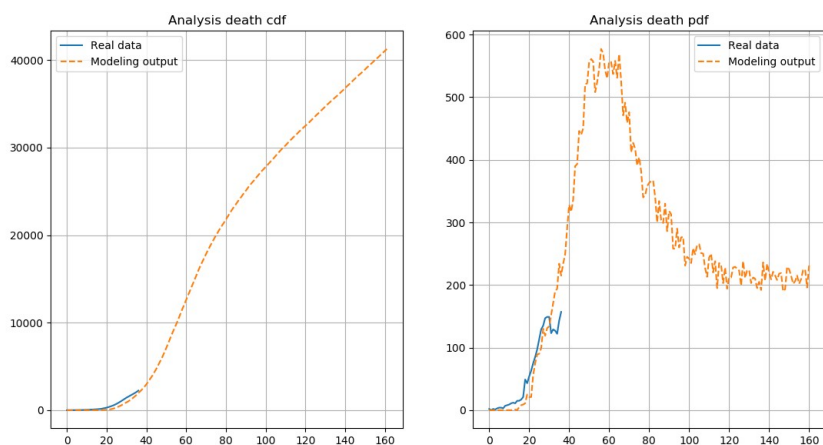
مجموع بهبودی‌ها: ۱۴۷۱۳۴۸ نفر

مجموع افراد در قرنطینه: ۶۲۳۹۵۵ نفر

همانطور که مشاهده می‌کنید نمودارهای مرگ و بیماری به طرز عجیبی درست پیش‌بینی می‌شود و حتی برخی نوسانات نویزی داده‌ها را پیش‌بینی می‌کند. اما در مورد بهبودی‌ها این روند کمی فرقی می‌کند و خطای زیادی را شاهد هستیم هر چند باید اذعان داشت با توجه به وجود قرنطینه‌های خانگی، افراد مشکوک و از همه مهمتر آمار به شدت نویزی از افراد بهبود یافته کلا نمودارهای مربوط به این افراد چندان قابل اعتماد نیست.

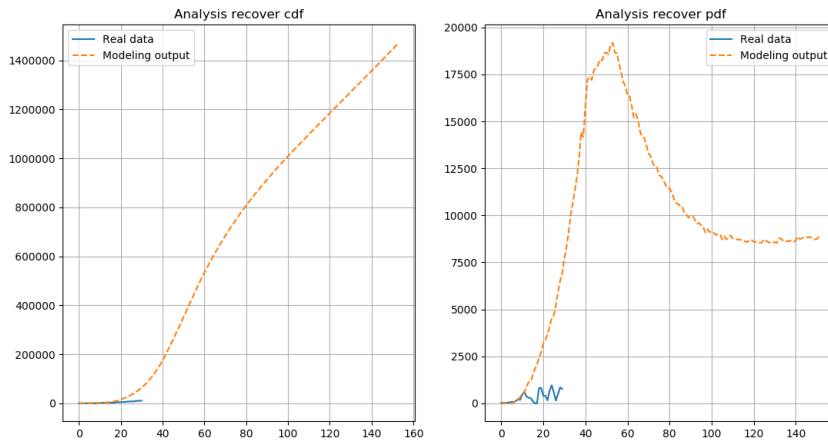


(نسخه جدید) نمودار ۱ - مقایسه مدل‌سازی و آمار تخمینی تعداد بیماران.



(نسخه جدید) نمودار ۲ - مقایسه مدل‌سازی و آمار فوتی‌هایی که رسماً تحت عنوان کرونا طبقه‌بندی شده‌اند^۱.

1 به دلیل صرفه جویی در کیت‌های تست کرونا، افرادی که پیش از تست کرونا فوت می‌کنند ولی علائم بالینی آن‌ها مشابه کروناست، بدون تست کرونا اعلام می‌شود. لذا آمار واقعی مرگ و میر ناشی از کرونا کمتر از آمار رسمی است.



(نسخه جدید) نمودار ۳- مقایسه مدلسازی و آمار قرنطینه.

تذکر:

در تحلیل خطای نمودارها سناریوهای زیر را در نظر بگیرید:

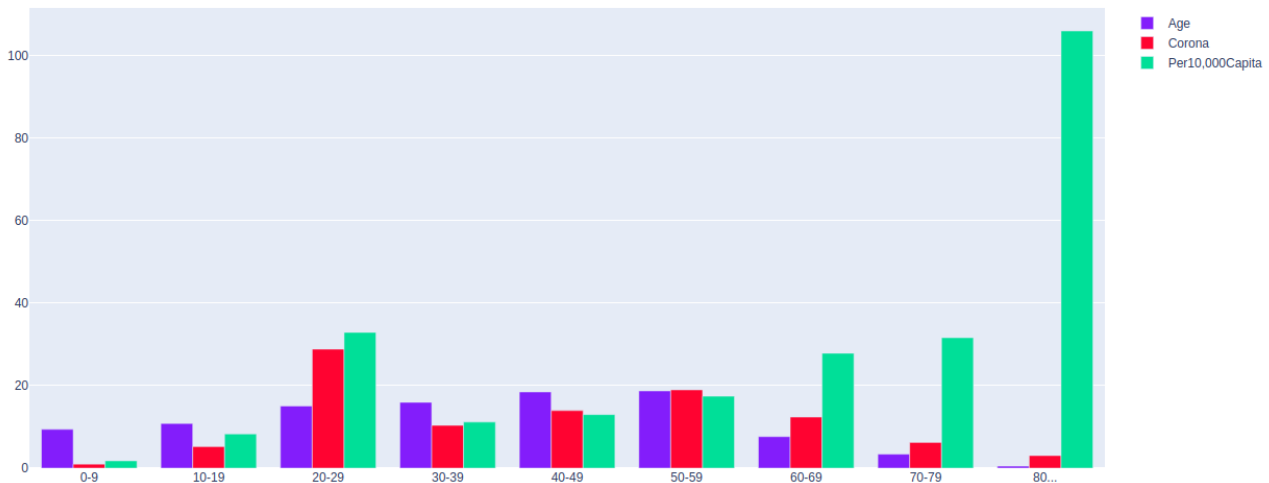
- (محتمل) خطای مدلسازی
- (محتمل‌ترین گزینه) تغییر و بهبود رسمی سیاست‌های دولتی و رفتار مردمی در کنترل بیماری که در مدلسازی لحاظ نشده است. در صورتی که این روند تا تاریخ ۲۹ اسفند تداوم داشته باشد، نیازمند ارائه یک نسخه جدید از مدلسازی و ارائه بهبود سیاست‌ها خواهیم داشت. قرنطینه خودخواسته مشکین شهر، عدم هجوم مجدد مردم به شمال، قرنطینه‌های اجباری محلی، افزایش آگاهی عمومی و بحران ایتالیا همگی در تغییر سیاست‌ها مؤثر بوده است.
- (غیرقابل بررسی) عدم ارائه آمار دقیق رسمی به منظور کنترل جو روانی جامعه. پیروی از چنین سیاستی می‌تواند به دو کارکرد مختلف بیانجامد. در صورتی که فرض حاکمیتی چنین سیاستی باشد باید این روند با ظرافت پیگیری شود. (قابل توجه افرادی که شبانه‌روز به دنبال آمار غیررسمی هستند، دوست عزیز تو بهداشتت رو رعایت کن، آمار به چه درد تو می‌خوره!):
 - بی‌اهمیت شدن بیماری برای مردم
 - بهداشت روانی مردم

پیوست

تحلیل ساختار جمعیتی

کره جنوبی

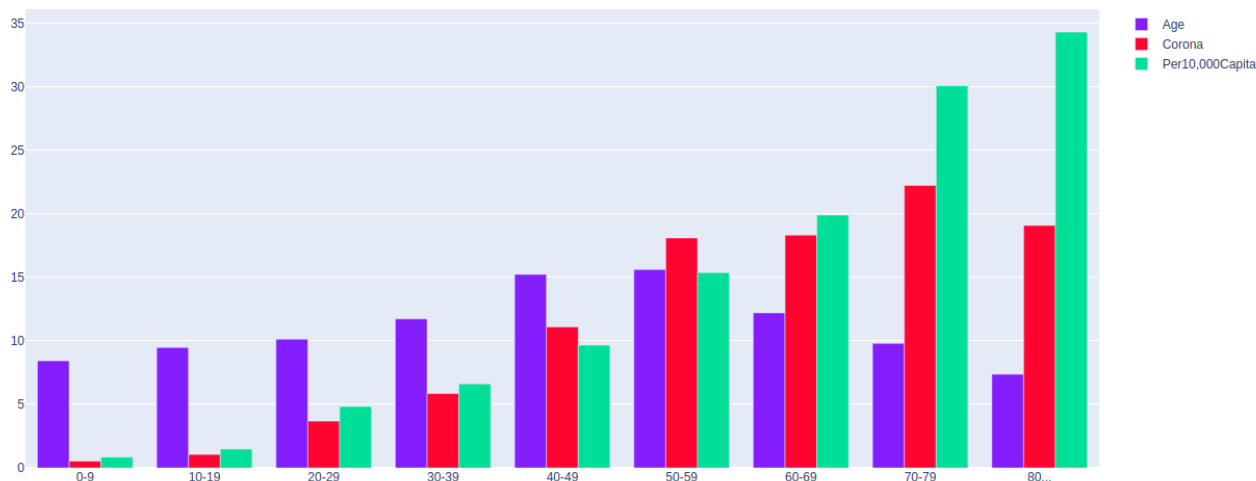
کره جنوبی در اقدامی تحسین برانگیز از جمعیت قابل توجهی از ساکنین بخش جنوبی شبه‌جزیره کره جنوبی توانست آمار خوبی از پراکندگی این ویروس ارائه کند. هیستوگرام سنی مردم کره، بیماران کره‌ای و بیمار در هر دهک را در نمودارهای زیر مشاهده می‌کنید:²



نمودار پیوست ۱ - مقایسه هرم سنی مردم و بیماران کره جنوبی و سرانه ابتلا به بیماری در بازه‌های سنی مختلف

ایتالیا

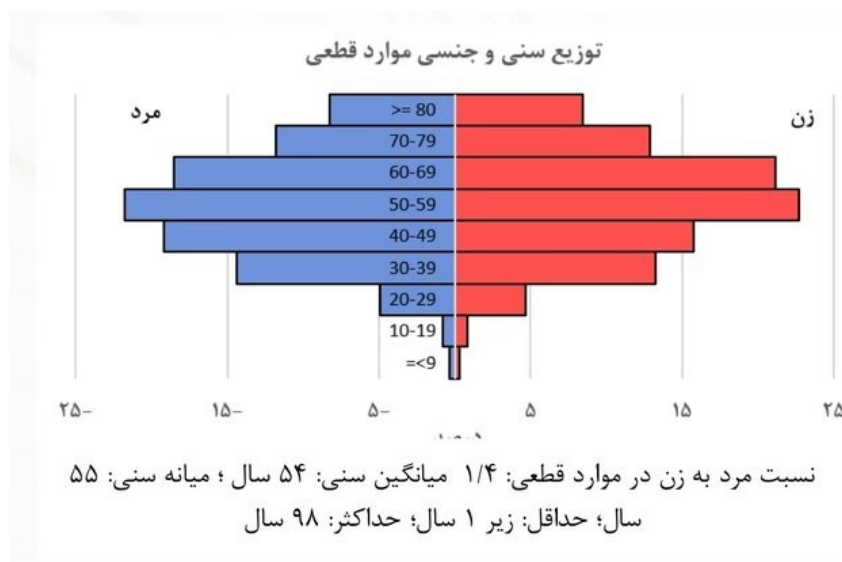
نه تنها ایتالیا در مهار بیماری با عمل‌کردی بسیار ضعیف و ناتوانی در فرهنگ‌سازی نیز نتوانست کاری از پیش ببرد بلکه در غربالگری کرونا نیز ضعیف عمل می‌کند. اما با توجه به در دسترس بودن ساختار سنی مبتلایان می‌توان در مورد جمعیت واقعی مبتلایان ایتالیا نظر داد. البته جدول سنی مبتلایان در ایتالیا متعلق به حدود ۵ روز پیش است که مجموع مبتلایان ایتالیا ۵ رقمی نشده بود.



نمودار پیوست ۲ - مقایسه هرم سنی مردم و بیماران رسمی ایتالیا و سرانه ابتلا به بیماری در بازه‌های سنی مختلف

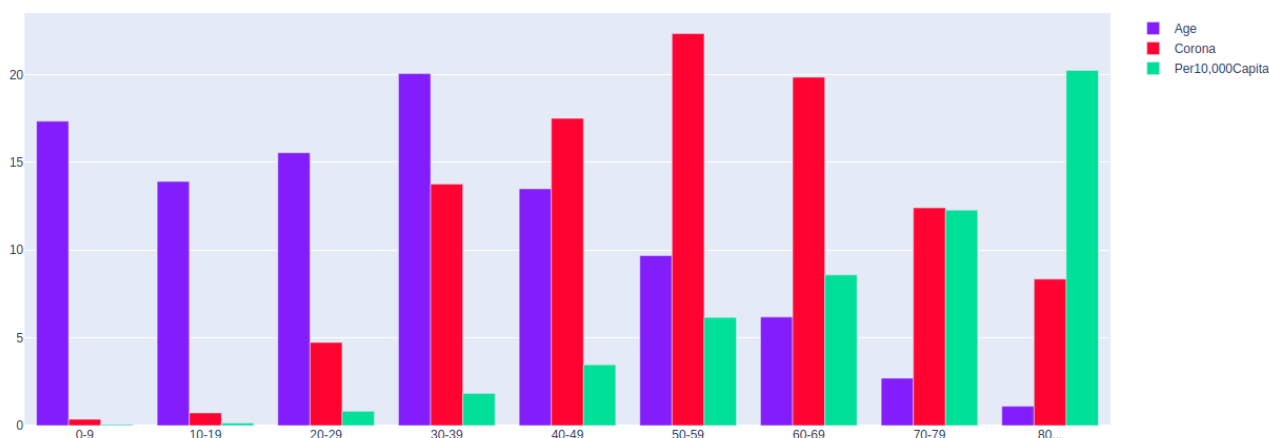
ایران

داده‌های ایرانی به صورت مجزا در دسترس نبود و در نموداری هم که آقای کیانوش جهانپور منتشر کرده‌اند³ نیز هیچ عدد و رقمی به چشم نمی‌خورد که قابل استفاده باشد. متأسفانه نه فقط عدم رسمیت توئیتر بلکه عدم احراز هویت اکانت آقای جهانپور باعث می‌شود نسبت به این آمار محتاط‌تر باشیم چرا که هر لحظه ممکنه است به هر دلیلی مالکیت آقای جهانپور بر این اکانت تکذیب شده و تمام اطلاعات آن باطل اعلام شود!



نمودار پیوست ۳ - هرم سنی بیماری در ایران

متأسفانه به دلیل انتشار بی‌سابقه اطلاعات فاقد عدد و رقم، به ناچار و کمک خط‌کش به استخراج اطلاعات پرداختیم.

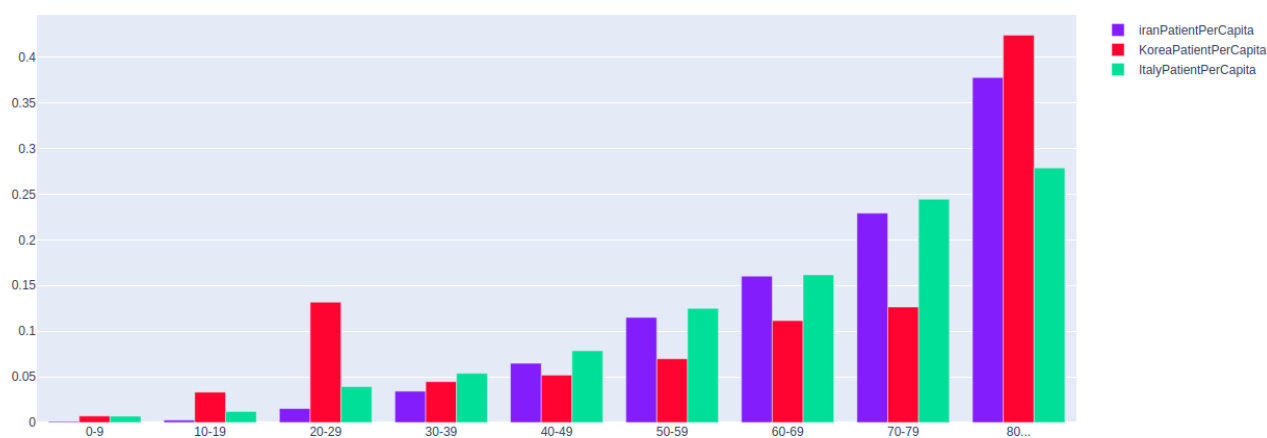


نمودار پیوست ۴ - مقایسه هرم سنی مردم و بیماران رسمی ایران و سرانه ابتلا به بیماری در بازه‌های سنی مختلف

مقایسه ایتالیا و کره جنوبی و ایران

با نرمال‌سازی نمودارها به^۴ خوبی تفاوتها و شباهتهای سه کشور در هرم سنی کرونا را مشاهده می‌کنید.

با توجه به اینکه کره از همه مردم تست کرونا گرفت، مرجع داده‌ها، داده‌های کره است. جهشی در دهک سوم می‌بیند که در دو کشور دیگر وجود ندارد، می‌تواند به هر دلیلی رخ داده باشد ولی با توجه به اینکه بدن جوانان در برابر این بیماری مقاومت بیشتری از خود نشان می‌دهد این بیماری در جوانان بروز کمتری دارد. لذا در ایران و ایتالیا که صرفاً از افراد دارای علائم تست می‌گیرند شاهد این جهش نیستیم.



نمودار پیوست ۵ - مقایسه هرم سنی بیماری در سه کشور ایران، کره و ایتالیا

هر چند ممکن است تفاوت‌های نژادی یا فرهنگی باعث تفاوت‌هایی در نمودار شوند. ولی با توجه به اینکه در علم آمار هرگاه اطلاعاتی از توزیع فضای نمونه نداشته باشیم توزیع را یکنواخت فرض می‌کنیم اینجا نیز مستثنی نیست، چون توزیع ساختار سنی بیماری را در کشورهای مختلف به ازای تفاوت‌های ژنتیکی و جغرافیایی مختلف را نمی‌دانیم در وهله نخست همه کشورها را یکسان فرض می‌کنیم. به مرور زمان و گسترش تست‌ها فرض ابتدایی را بهبود می‌دهیم ولی مفروض این است که ایران و ایتالیا نیز مانند کره جنوبی هستند.

با توجه به نمودار کره، سعی می‌کنیم نسبت بین همه بیماران و بیماران شناسایی شده در دو کشور را پیدا کنیم.

در ایران و ایتالیا در دهکهای زیر ۳۰ سال و بالاترین دهک شاهد اشکال اساسی هستیم. یعنی آمار رسمی کمتر از بیماران واقعی است (نمودار کره مرجع می‌باشد). چهار دهک مشکل دار در کره حدود ۶۶٪ همه داده‌ها را تشکیل می‌دهند در حالیکه در ایران ۶۰٪ و ایتالیا ۳۴٪ از بیماران را تشکیل می‌دهند. با روند زیر به دنبال پیدا کردن تعداد افراد بیماری هستیم که هنوز کشف نشده‌اند.

• مفروض این است که در دهک‌هایی که آمار کره کمتر از آمار کشور هدف است غربالگری درست انجام شده باشد (فرض ساده‌سازی خوش‌بینانه) ولی همچنان حدود ۸۰٪ بیماران ناشناخته باشند.

• تشکیل بردار

$$y = [\text{symbol}(x), \text{sum of real decades} * 5]$$

(دلیل ضرب در چهار شدن قسمت دوم این است که تنها یک پنجم بیماران واقعی علائم بالینی دارند و در کشورهای ایران و ایتالیا تست می‌شوند).

$$\text{تشکیل تناسب} \quad \frac{x}{y} = \frac{06}{04} \quad \text{یعنی} \quad \frac{x}{y} = 1.5$$

• جایگذاری $y = 1$

• حال در صورتی که بخواهیم هرم سنی مناسب برای ایران را رسم کنی، x را باید به نسبت مقدار ستون‌های مربوط به زیر ۳۰ سال و بالای ۸۰ سال تقسیم کنیم و به هر ستون عددی جدید نسبت بدهیم (فعالاً مد نظر ما نیست).

• با توجه به اطلاعات فوق، برای واقعی کردن نمودار سنی بیماران ایران نیازمند ۶۰٪ بیمار بیشتر هستیم و ایتالیا ۵۰٪ بیشتر. (یعنی این حداقل این تعداد بیمار ناشناخته داریم و باید به این تعداد مریض به مجموع مریض‌ها اضافه کنیم.) به این عدد نسبت سنی کشور می‌گوییم.

نتیجه‌گیری

می‌دانیم تعداد واقعی بیمار ضریبی از هرم سنی خواهد بود. از طرف دیگر می‌دانیم تنها ۲۰٪ از ناقلین بیمار هستند که نزدیک به نصف آن‌ها نیز به دلیل آلوده بودن بیمارستان‌ها و وخیم نبودن وضعیت بالینی از حضور در بیمارستان خودداری می‌کنند. با این اوصاف به دنبال ضریبی از نسبت سنی کشور می‌گردیم که بین ۵ تا ۹ باشد. به نظر می‌رسد اعداد ۸/۲ و ۷/۵ که ۵ برابر ضریب سنی ایران و ایتالیا هستند واقع‌گرایانه‌ترین گزینه‌ها باشند. زیرا در ایران سیاست قرنطینه خانگی بیماران بدون زمینه‌ی خانگی و تبلیغات بر عدم حضور در مراکز درمانی و در ایتالیا سیاست عدم تست گرفتن از افراد دارای بیماری زمینه‌ی و سالخورده ما را به انتخاب اعدادی راهنمایی می‌کند به ۹ نزدیک‌تر باشند تا ۵. (دلیل عدم انتخاب ۹ برای ایتالیا صرفاً اکراه نویسنده از سیاست بسیار بدبینانه بود.).

به طور خلاصه برای پیدا کردن تعداد ناقلین ایرانی باید آمار رسمی در ۸/۲ ضرب شود. برای پیدا کردن تعداد ناقلین ایتالیایی، آمار رسمی باید در ۷/۵ ضرب شود.

ایتالیا	ایران	
۱۴۱۷۷۷	۱۳۲۵۸۵	ناقل
۳۱۵۰۶	۱۶۱۶۹	بیمار

جدول پیوست ۱ - آمار رسمی و تخمینی تا صبح روز ۲۸ اسفند (پیش از ارائه رسمی روزانه)