

## Corona Virüs Salgını Teşhis Sayısı Öngörü Analizi

Arif Selçuk Öğrenci

29.3.2020, 14:00

0. Aşağıdaki tüm analizler temelde dinamik sistemin verisinin matematik modellemesine dayanmaktadır. Hastalık yayılımı ile ilgili bir tıbbi modelleme veya bir tıbbi/biyolojik uzmanlık barındırmamaktadır. Amaç, Türkiye'nin ilk 18 güne ilişkin verisine dayanarak toplam vaka sayısını öngörmenin ne kadar olanaklı olduğunu bulmaktır. Bunun için doyuma ulaşmış ve hala artışta olan bazı ülkelerin verileri incelenecektir. Bu kapsamda bu ülkeler için verinin tamamı yerine ilk-"n" günü kullanılarak farklı sig-3 regresyon modelleri elde edilmiş, burada ulaşılması beklenen maksimum hasta sayısının bir desene sahip olup olmadığı, günlük değişimlerin nasıl bir oynaklığa yol açtığı incelenmiştir.

Bu analizin 5 günde bir güncellenmesi planlanmaktadır.

1. Kullanılan veri kaynağı: <https://ourworldindata.org/coronavirus-source-data>
2. Modelde kullanılan iki fonksiyon  
3-parametrelili sigmoid (sig-3)

$$y = \frac{a}{1 + e^{-\frac{x-x_0}{b}}}$$

3-parametrelili lojistik (log-3)

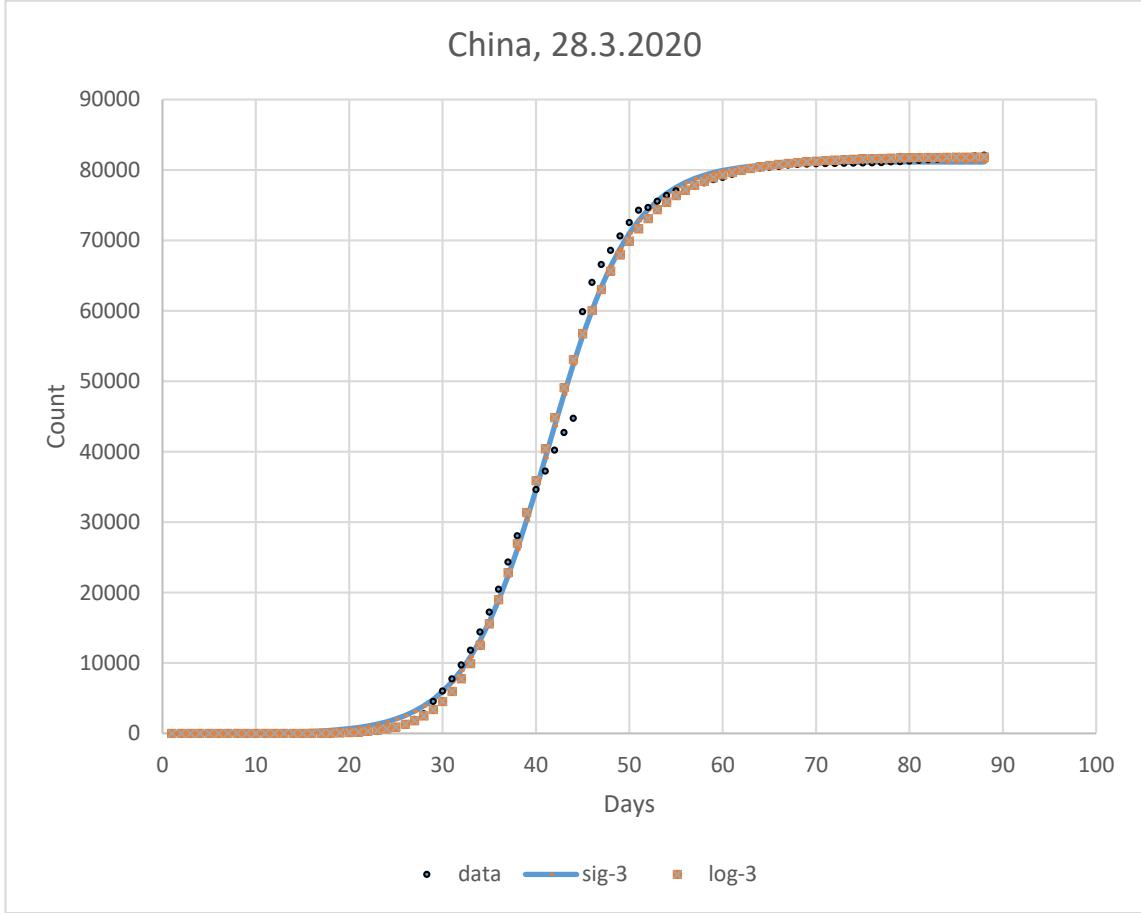
$$y = \frac{a}{1 + \left(\frac{x}{x_0}\right)^b}$$

Bu fonksiyonlar salgın modellemesinde SIR, SEIR modellerinde karşımıza sıklıkla çıkmaktadır. Pratik olarak, sig-3 modelinde "a" parametresi en sonunda oluşacak toplam enfekte/hasta sayısını vermektedir. "x0" parametresi oluşan eğrinin büküm noktasını (eğiminin maksimum olduğu nokta), "b" parametresi de eğrinin dikliğini göstermektedir.

R<sup>2</sup> değeri en basit anlatımla yapılan regresyonun veri noktalarından "ne kadar iyi" geçtiğinin göstergesidir, bu değer 0-1 aralığında olur ve 1 olması, eğrinin veri noktalarının tam üstünden geçtiğini, mükemmel bir "fit" olduğunu gösterir. Sonuçlarda görülen 0.99xxx gibi bir değer oldukça başarılı bir fit yapıldığını gösterir.

3. Aşağıda farklı ülkelerin enfekte hasta sayısı verisine bu iki fonksiyonla SIGMAPLOT yazılımında regresyon uygulanarak model parametreleri elde edilmiştir. Yatay eksen ilk hastanın ilan edildiği andan itibaren gün sayısını (Days), dikey eksen de hasta sayısını (Count) vermektedir.

4. ÇİN  
88 günlük veri



ÇİN	sig-3	log-3
a	81138.1	81923.86
b	4.462855	-9.00711
x0	41.31811	41.12196
R <sup>2</sup>	0.998362	0.997857

Yorum: Her iki fonksiyon da oldukça yakın bir şekilde veriyi modellemektedir. 80 gün civarında doyuma ulaşmış görünen eğri için regresyon modelleri 41 günde bir büküm noktası öngörmektedir. Ulaşılan maksimum hasta değeri 82079 olup, regresyon modelleri bu değere %1.1 yakınlıktadır.

ÇİN verisinin tamamı yerine ilk-n günü kullanılarak sig-3 regresyon modelleri elde edilmiş ve bunların doyum değerleri ("a" parametresi) aşağıda verilmiştir.

Kullanılan gün sayısı	a
20	(*)
25	(*)
30	48910
35	26390
40	48120
45	71020

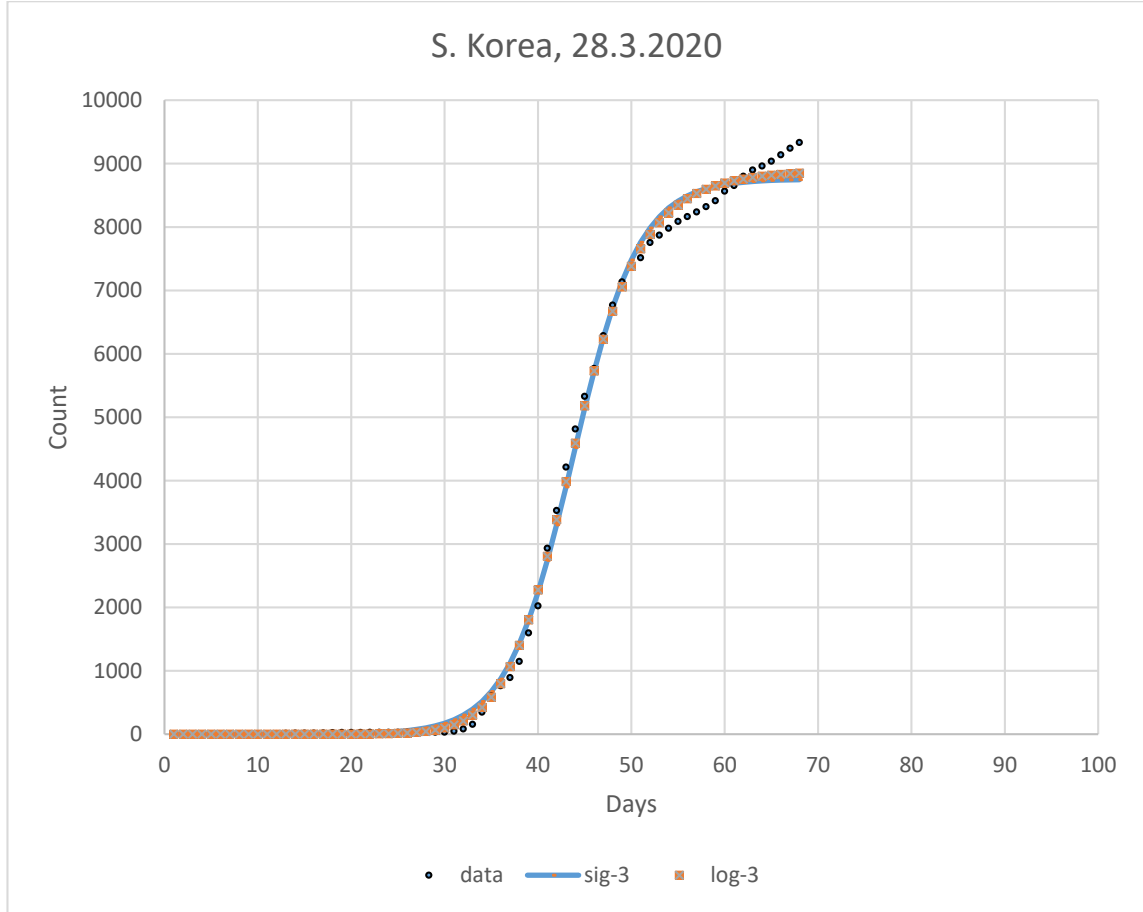
50	90190
55	82860
60	80990
65	80830
70	80860
75	80900
80	80950
88	81138

(\*) 20 ve 25 gün için yapılan regresyonda algoritma yakınsamayı başaramamıştır.

Yorum: İlk 30 günü kullanarak regresyon yapıldığında maksimum değer 48910 çıkmaktadır, ancak sistemin gerçek dinamiği ancak ilk 45 günden sonra tam olarak anlaşılmaktadır. Burada ilk 45 günde yapılan regresyonun 88 güne göre hatası (81138-71020) %12 civarındadır.

##### 5. G. KORE

68 günlük veri



G. KORE	sig-3	log-3
a	8756.041	8897.114
b	3.524001	-11.8825
x0	43.76948	43.76377
R <sup>2</sup>	0.997411	0.998371

Yorum: Veri doyuma ulaşmamıştır, 50 günden sonra doğrusal bir değişim gözlenmiştir. Ancak sigmoidal bölgede her iki fonksiyon da oldukça yakın bir şekilde veriyi modellemektedir. 80 gün civarında doyuma ulaşacak regresyon modelleri 44 günde bir büküm noktası öngörmektedir.

G. Kore verisinin tamamı yerine ilk-n günü kullanılarak sig-3 regresyon modelleri elde edilmiş ve bunların doyum değerleri ("a" parametresi) aşağıda verilmiştir.

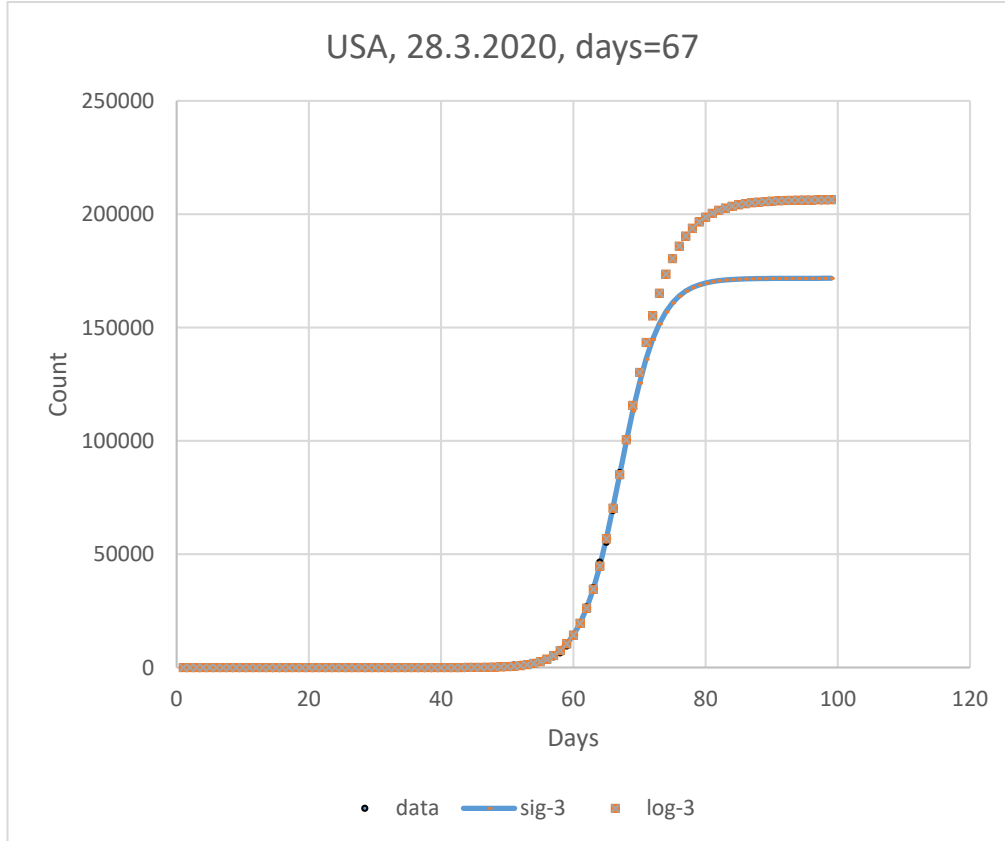
Kullanılan gün sayısı	a
20	30
25	30
30	30
35	51440(*)
40	4077
45	7141
50	7858
55	8117
60	8315
68	8756

(\*) 35 gün için yapılan regresyonda algoritma yakınsamayı başaramamıştır.

Yorum: İlk 20 günü kullanarak regresyon yapıldığında maksimum değer 30 çıkmaktadır, ancak sistemin gerçek dinamği ancak ilk 45 günden sonra tam olarak anlaşılmaktadır. Burada ilk 45 günde yapılan regresyonun 68 güne göre hatası (8756-7141) %20 civarındadır.

## 6. USA

67 günlük veri



USA	sig-3	log-3
a	171721.5	206513.3
b	2.939834	-20.2975
x0	67.06216	68.18214
R^2	0.999285	0.999337

Yorum: Veri doyuma ulaşmamıştır, HALA sigmoidal bölgede olup her iki fonksiyon da oldukça yakın bir şekilde veriyi modellemektedir. 90 gün civarında doyuma ulaşacak regresyon modelleri 67 günde bir büküm noktası öngörmektedir. Log-3 modeli daha yüksek bir maksimum öngörmektedir.

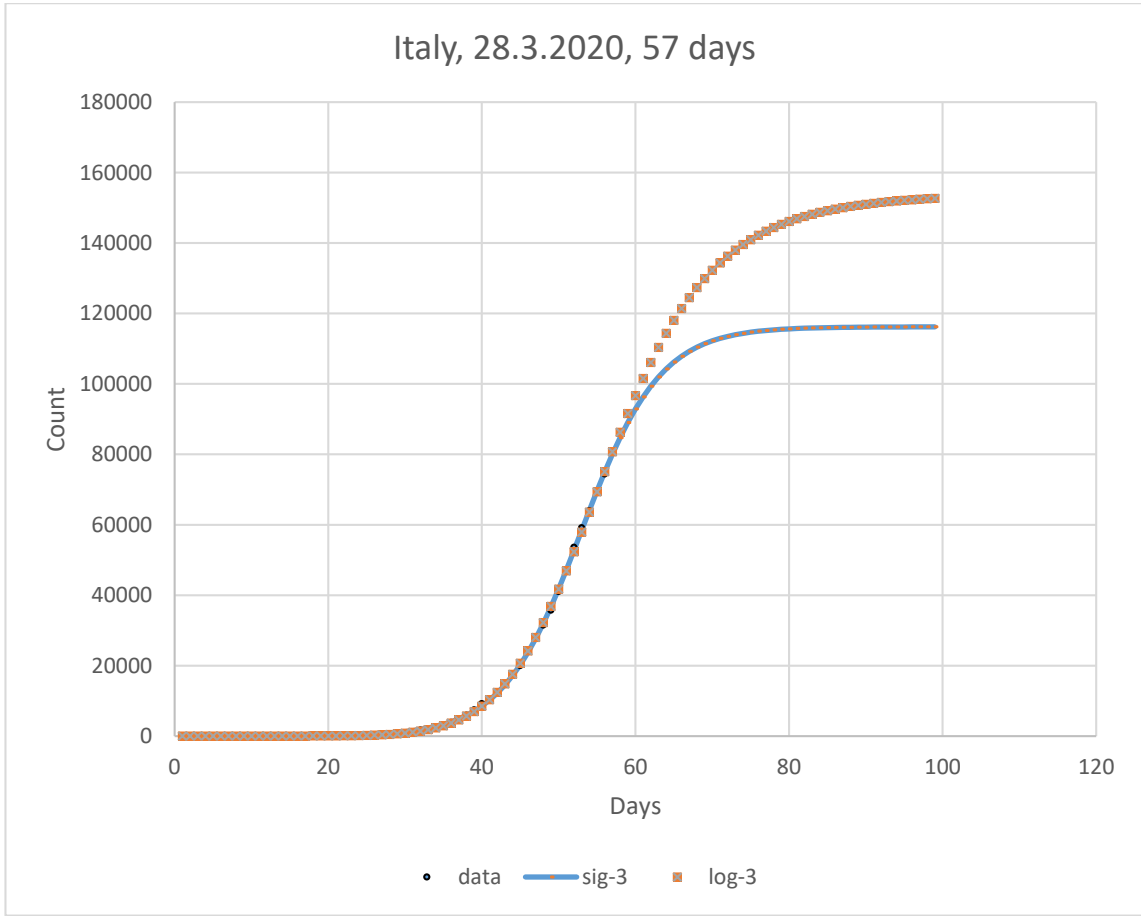
USA verisinin tamamı yerine ilk-n günü kullanılarak sig-3 regresyon modelleri elde edilmiş ve bunların doyum değerleri ("a" parametresi) aşağıda verilmiştir.

Kullanılan gün sayısı	a
20	
25	
30	15
35	36790
40	100300
45	232500
50	668000
55	675900
60	(*)
65	108600
67	171721

(\*) 60 gün için yapılan regresyonda algoritma yakınsamayı başaramamıştır.

Yorum: İlk 30 günde hasta sayısı 15'dir. Sistemin gerçek dinamiği ortaya çıkmamıştır ancak ilk 30+45 =75 günden sonra tam olarak anlaşılması beklenebilir. Birkaç günlük eklemeler bile tahminlerde ciddi farklara yol açmaktadır.

7. İTALYA  
57 GÜNLÜK VERİ



Italy	sig-3	log-3
a	116205.9	154095.5
b	5.103287	-8.29438
x0	52.95111	56.34824
R <sup>2</sup>	0.999727	0.999719

Yorum: Veri doyuma ulaşmamıştır, HALA sigmoidal bölgede olup her iki fonksiyon da oldukça yakın bir şekilde veriyi modellemektedir. 90 gün civarında doyuma ulaşacak regresyon modelleri 53 günde bir büküm noktası öngörmektedir. Log-3 modeli daha yüksek bir maksimum öngörmektedir.

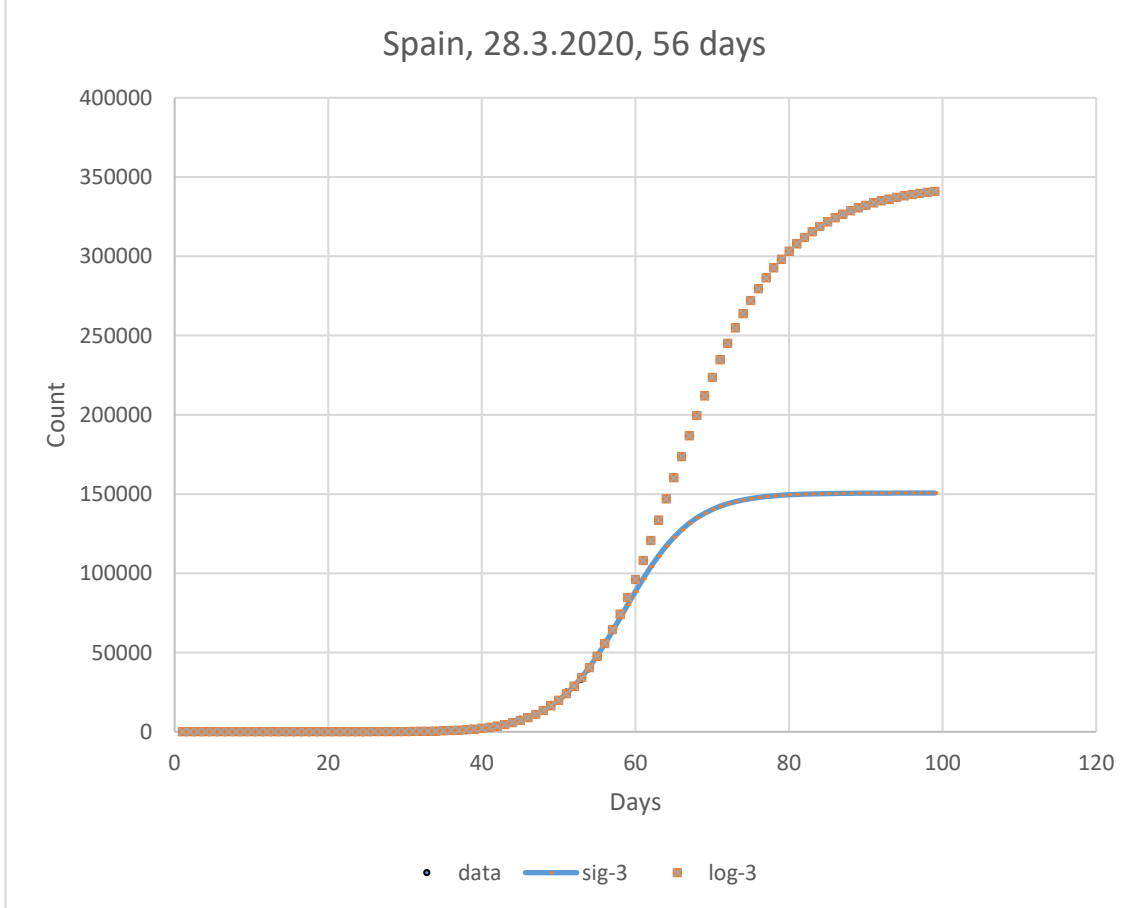
İtalya verisinin tamamı yerine ilk-n günü kullanılarak sig-3 regresyon modelleri elde edilmiş ve bunların doyum değerleri ("a" parametresi) aşağıda verilmiştir.

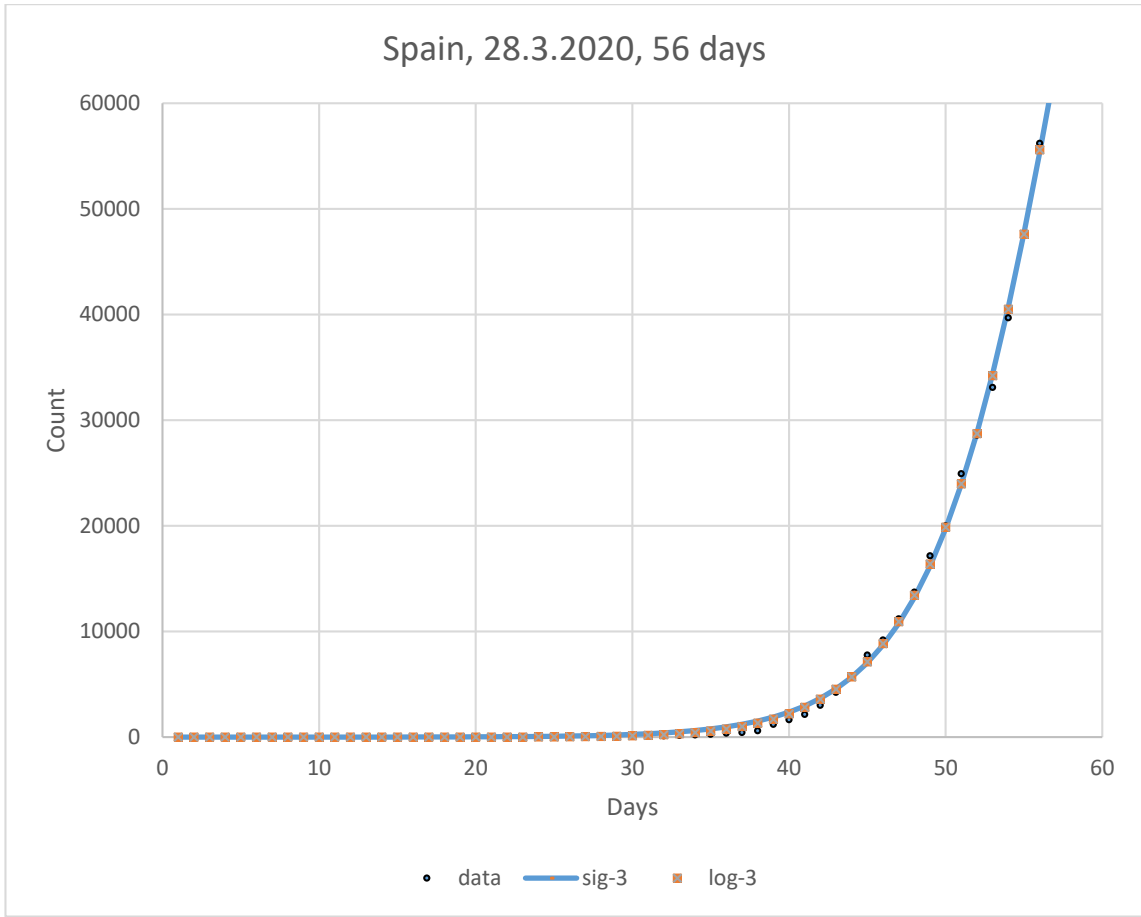
Kullanılan gün sayısı	a
20	
25	144
30	2294
35	5976
40	38850
45	42500

50	83990
55	114900
57	116205

YORUM İtalya verisi çok düzgün olmamakla birlikte, en azından anormal zıplamalar yapmadan, maksimum değeri artırmaktadır. Verinin 65 (20. Güne kadar düşük olan sayı + 45 günlük büküm, 20+45=65) günden itibaren “doğru” maksimumu vermesi beklenebilir.

8. İSPANYA  
56 GÜNLÜK VERİ





SPAIN	sig-3	log-3
a	150675.1	346608.4
b	4.45008	-10.0974
x0	58.42115	65.97722
R <sup>2</sup>	0.998683	0.999182

Yorum: Veri doyuma ulaşmamıştır, HALA sigmoidal bölgede olup her iki fonksiyon da oldukça yakın bir şekilde veriyi modellemektedir. 90 gün civarında doyuma ulaşacak regresyon modelleri 58 günde bir büküm noktası öngörmektedir. Log-3 modeli daha yüksek bir maksimum öngörmektedir.

İspanya verisinin tamamı yerine ilk-n günü kullanılarak sig-3 regresyon modelleri elde edilmiş ve bunların doyum değerleri ("a" parametresi) aşağıda verilmiştir.

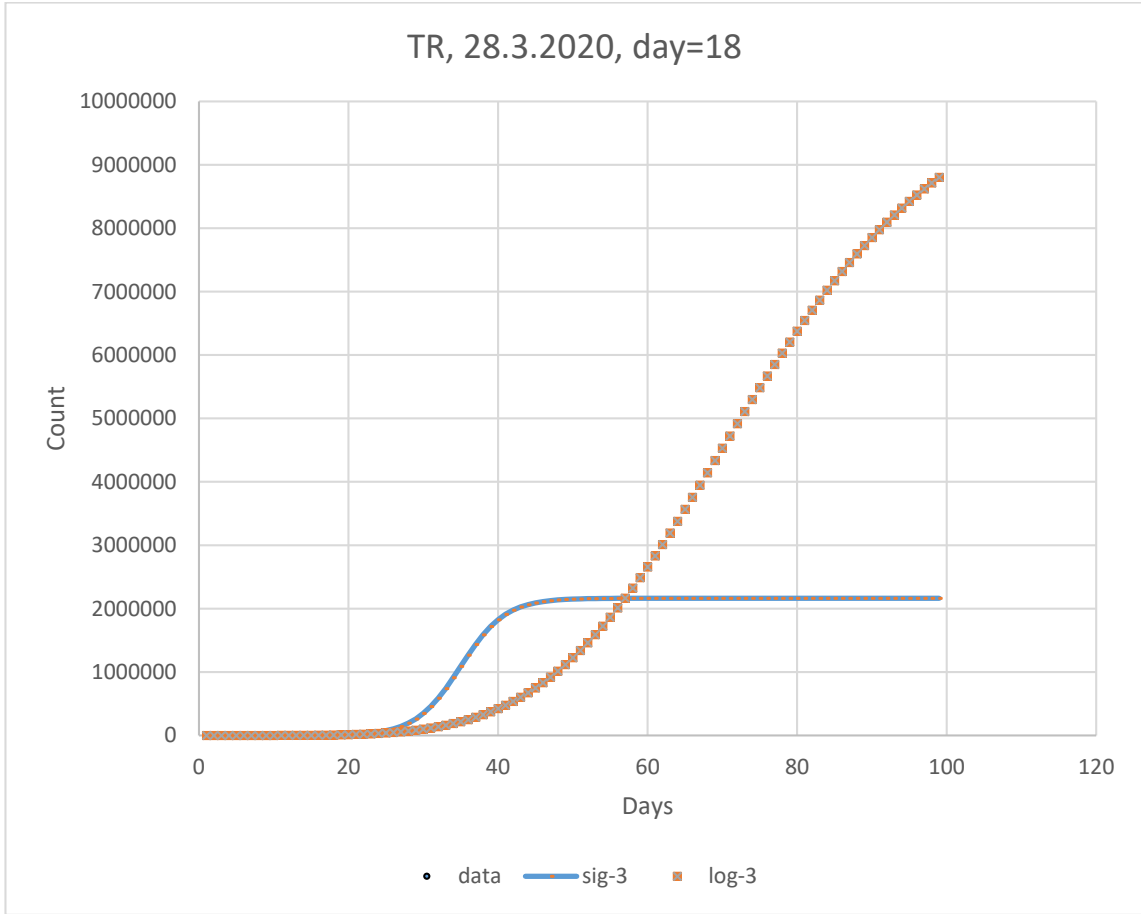
Kullanılan gün sayısı	a
20	
25	
30	37270
35	527

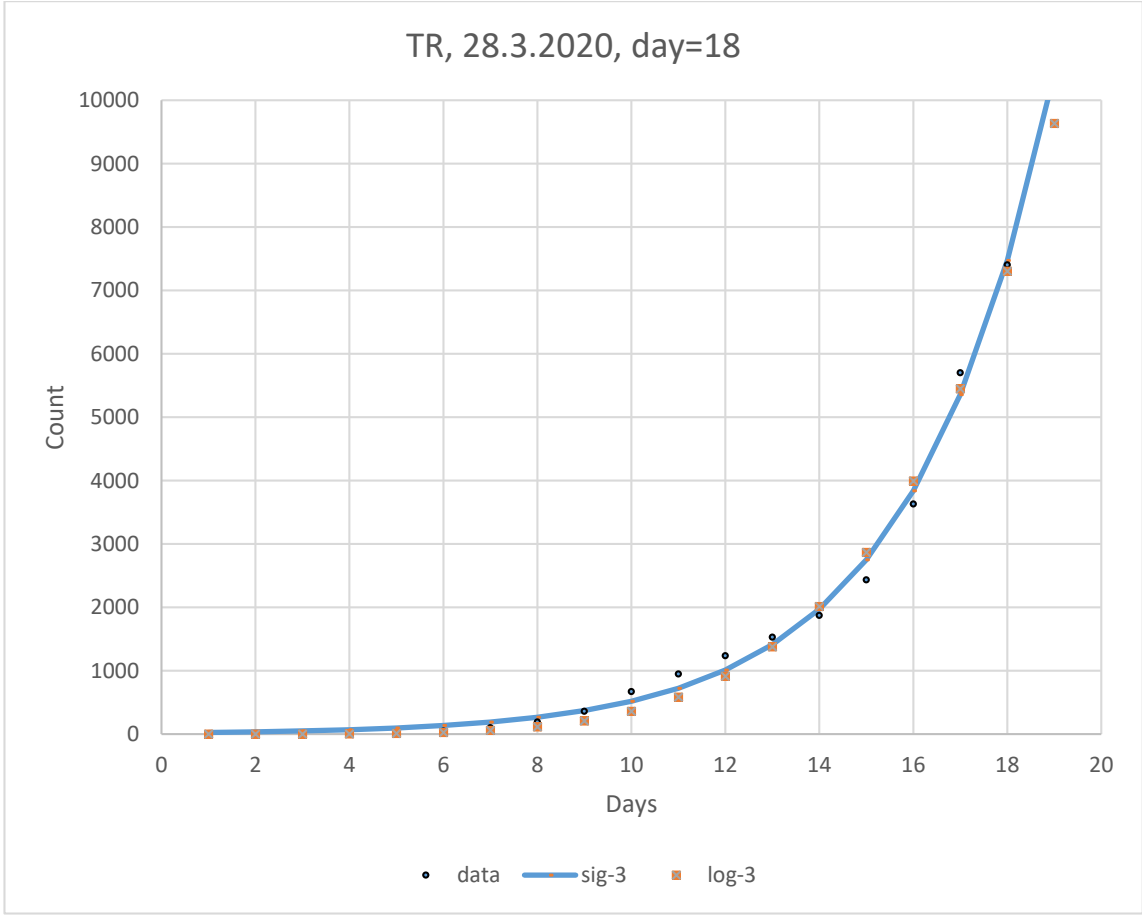


40	1384000
45	34990
50	35820
55	110100
56	150675

YORUM İspanya verisi anormal zıplamalar yapmaktadır. Verinin 70 (25. güne kadar düşük olan sayı + 45 günlük büküm, 25+45=70) günden itibaren "doğru" maksimumu vermesi beklenebilir.

## 9. TÜRKİYE 18 GÜNLÜK VERİ





TR	sig-3	log-3
a	2164010	10893223
b	2.994145	-5.13025
x0	34.96183	74.79085
R <sup>2</sup>	0.994184	0.989646

Yorum: Veri doyuma ulaşmamıştır, HALA sigmoidal bölgede olup her iki fonksiyon da oldukça yakın bir şekilde veriyi modellemektedir. SİG-3 İÇİN 50, LOG-3 İÇİN 150 gün civarında doyuma ulaşacak regresyon modelleri 35/75 günde bir büküm noktası öngörmektedir. Log-3 modeli daha yüksek bir maksimum öngörmektedir.

TR verisinin tamamı yerine ilk-n günü kullanılarak sig-3 regresyon modelleri elde edilmiş ve bunların doyum değerleri ("a" parametresi) aşağıda verilmiştir.

Kullanılan gün sayısı	a
10	3912
11	1443
12	1596

13	1850
14	2241
15	3320
16	24930
17	5536952
18	2164010

YORUM TR verisi anormal zıplamalar yapmaktadır. Verinin35 (sayı YÜKSELİŞE HEMEN BAŞLAMIŞTIR + 35 günlük büküm, 0+35=35) günden itibaren “doğru” maksimumu vermesi beklenebilir.

#### GENEL YORUM

Sistem dinamikleri deęişkendir, uygulanan yaklaşımlar ve veri yayınlama-analiz politikaları farklıdır, sağlıklı bir öngörünün yapılabilmesi için

- Verinin anlamlı yükselişe geçmesinin üzerinden 35 gün geçmesi gerekli görülmektedir.

Dinamiğin (önlemler, tedavi) deęişmesi sebebiyle günlük deęişimlerin etkisi yüksektir, yanıltıcı olabilir. TR verisi için öngörünün sağlıklı olabilmesi için 17 gün daha beklenmelidir. Ancak şu anki üstel hareketin bir süre daha, tahminimce en az 2 hafta, devam etmesi de beklenir.