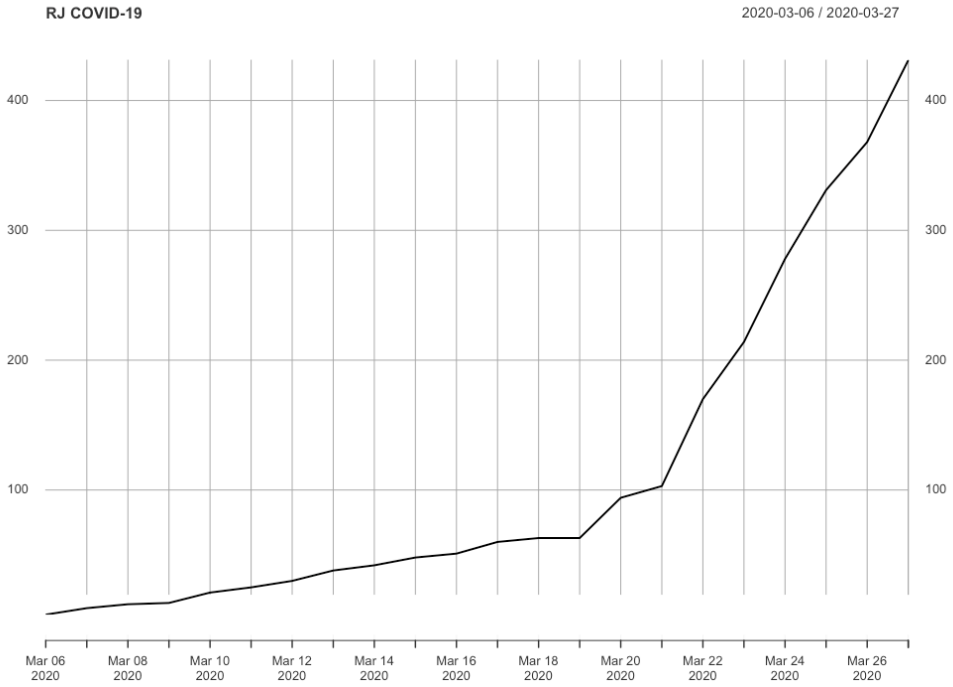
Previsão da evolução dos casos de COVID-19 no município do Rio de Janeiro para o período de 28/março a 03/abril

O LEGOS trata o problema de prever a evolução do total de casos confirmados de COVID-19 no município do Rio de Janeiro. Para chegar aos resultados que serão apresentados a seguir utilizamos a série histórica do referido município que está disponível no Painel Rio COVID SMS | IPP | COR[1]. Esta série histórica é apresentada na tabela e figura abaixo.

Data	Casos confirmados	Data	Casos confirmados
06/03/2020	4	17/03/2020	60
07/03/2020	9	18/03/2020	63
08/03/2020	12	19/03/2020	63
09/03/2020	13	20/03/2020	94
10/03/2020	21	21/03/2020	103
11/03/2020	25	22/03/2020	170
12/03/2020	30	23/03/2020	214
13/03/2020	38	24/03/2020	278
14/03/2020	42	25/03/2020	331
15/03/2020	48	26/03/2020	366
16/03/2020	51	27/03/2020	431

[1] Link: https://experience.arcgis.com/experience/38efc69787a346959c931568bd9e2cc4



Observando a figura acima podemos perceber que no município do Rio de Janeiro segue comportamento não linear. Entretanto, se separarmos este gráfico em 3 momentos diferentes podemos observar que:

1. Dia 07/03 ao 16/03 segue um "comportamento linear";
2. Dia 17/03 ao 19/03 varia de apenas 3 casos e nos dias 18 e 19 possuem o mesmo valor de 63 casos confirmados.
3. Dia 20/03 até o momento em que tivemos informação disponível (boletim de 27/03/2020 18:00) segue um "comportamento linear" novamente, mas bem maior que o observado no período de 07/03 ao 16/03.

Assim, para reproduzirmos o comportamento esperado para a próxima semana consideraremos que a base histórica útil à nossa série se inicia no dia 20/03 (quando o total de casos confirmados de pacientes com COVID-19 ultrapassa 90) e perdura até o momento. A alteração do "ponto de corte" da série em relação às notas técnicas anteriores levará a um resultado diferente por conta dos com menores resultados de erro encontrados serem nos modelos lineares.

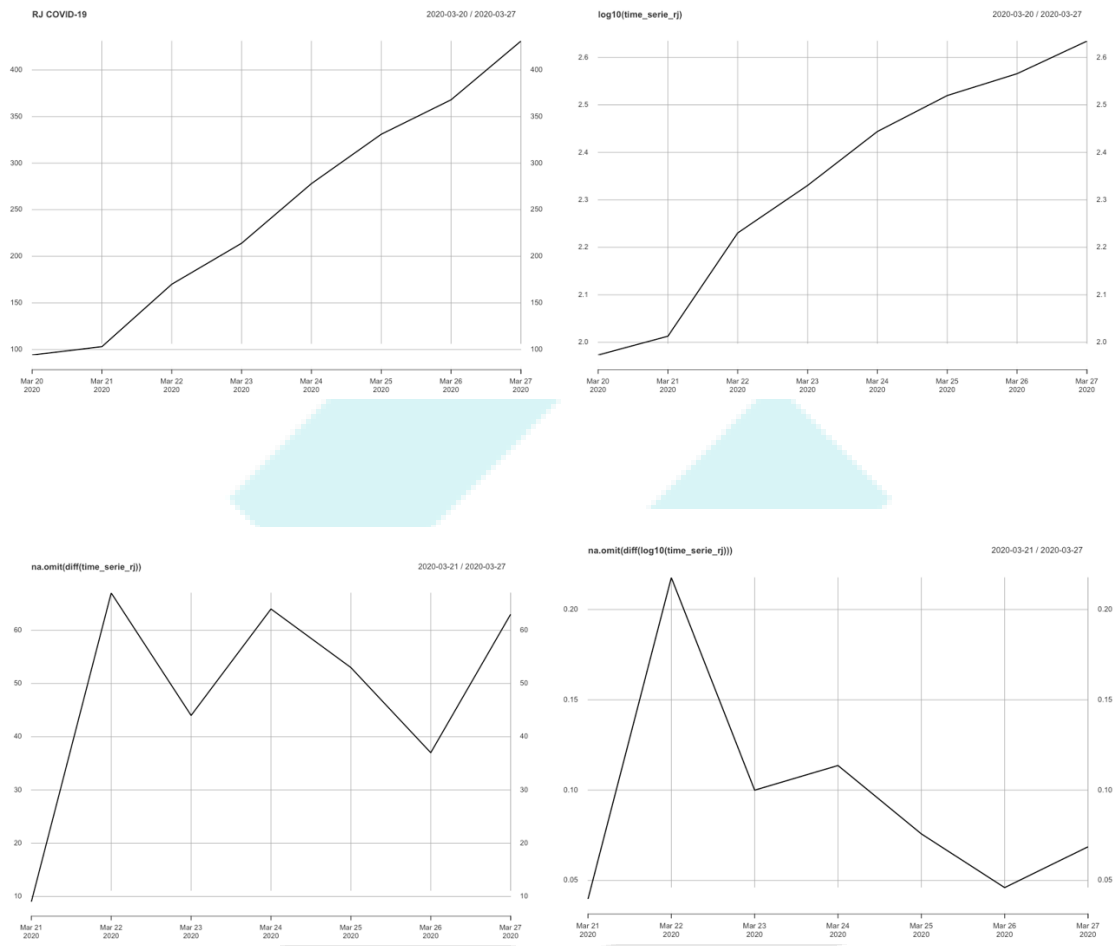
Modelos de previsão

Para estimarmos o total de casos de pacientes com COVID-19 confirmados no horizonte de 7 dias a frente (até 03/04) utilizamos modelos univariados consagrados da literatura e regressões apresentados abaixo:

1) Modelos de suavização exponencial (ETS);
2) Modelos auto regressivos integrados de média móvel (ARIMA);
3) Regressão Linear (RL);

Nenhum dos 3 modelos apresentados acima foi concebido para tratar de séries temporais de comportamento explosivo (como é o nosso caso). Entretanto, podemos realizar algumas transformações na série original apresentada para que a série transformada atenda às condições de "contorno" das técnicas apresentadas acima. Neste sentido, as

principais formas comumente utilizadas em séries temporais são: diferenciação e transformação logarítmica (elas também podem ser combinadas). Nas figuras abaixo são apresentadas a série original e cada série transformada (logarítmica, diferenciada e logarítmica-diferenciada) de casos confirmados no Município do Rio de Janeiro.



Como o objetivo de trabalhar sobre a série mais estacionária possível (ainda que com tendência) trabalharemos sobre as duas séries que melhor apresentam este comportamento. A partir deste ponto chamaremos as séries escolhidas para prever o total de casos confirmados de série original (SO) e a série "logaritimatizada" (SL) e apresentaremos abaixo características observadas de cada modelos e cada série.

1) ETS na série original (ETS SO): não capturou o padrão de crescimento e se manteve constante ao longo do tempo;
2) ETS na série logarimitizada (ETS SL): não capturou o padrão de crescimento e se manteve constante ao longo do tempo;
3) ARIMA na série original (ARIMA SO): capturou o padrão de crescimento linear;
4) ARIMA na série logarimitizada (ARIMA SL): capturou o padrão de crescimento explosivo;
5) Regressão linear na série logarimitizada (RL SO): capturou o padrão de crescimento linear;
6) Regressão linear na série logaritimizada (RL SL): capturou o padrão de crescimento explosivo;

As previsões otimista, esperada e pessimista de acordo com cada modelo são apresentados nas tabelas abaixo. O intervalo de confiança escolhido para a geração dos cenários otimista e pessimista foi de 95%, ou seja, a probabilidade de observarmos um valor fora desta faixa de valores é de 5% caso o "comportamento linear" da série se mantiver.

CENÁRIO OTIMISTA	ETS SO	ARIMA SO	RL SO	ETS SL	ARIMA SL	RL SL
28/03/2020	321	439	433	252	408	419
29/03/2020	276	470	481	202	453	514
30/03/2020	241	506	528	171	516	627
31/03/2020	212	543	574	148	596	763
01/04/2020	186	582	621	130	695	927
02/04/2020	163	621	667	116	815	1124
03/04/2020	141	662	713	105	960	1361

CENÁRIO ESPERADO	ETS SO	ARIMA SO	RL SO	ETS SL	ARIMA SL	RL SL
28/03/2020	431	479	475	431	536	613
29/03/2020	431	527	526	431	666	771
30/03/2020	431	575	576	431	828	971
31/03/2020	431	624	626	431	1029	1221
01/04/2020	431	672	677	431	1279	1536
02/04/2020	431	720	727	431	1590	1933
03/04/2020	431	768	777	431	1976	2432

CENÁRIO PESSIMISTA	ETS SO	ARIMA SO	RL SO	ETS SL	ARIMA SL	RL SL
28/03/2020	541	519	517	736	704	897
29/03/2020	586	584	571	919	979	1159
30/03/2020	621	645	624	1089	1328	1503
31/03/2020	650	704	678	1257	1776	1955
01/04/2020	676	762	733	1427	2354	2547
02/04/2020	699	818	787	1599	3102	3324
03/04/2020	721	874	842	1776	4067	4344

Conforme comentado anteriormente, podemos descartar a priori os modelos ETS SO e ETS SL porque não foram capazes de capturar o crescimento que devemos prever. Assim, analisaremos apenas os resultados dos modelos ARIMA SO, ETS SL, ARIMA SL e RL SL. Para selecionarmos qual modelo deveríamos usar para estimar a quantidade total de casos esperados para os próximos 7 dias utilizamos os seguintes critérios:

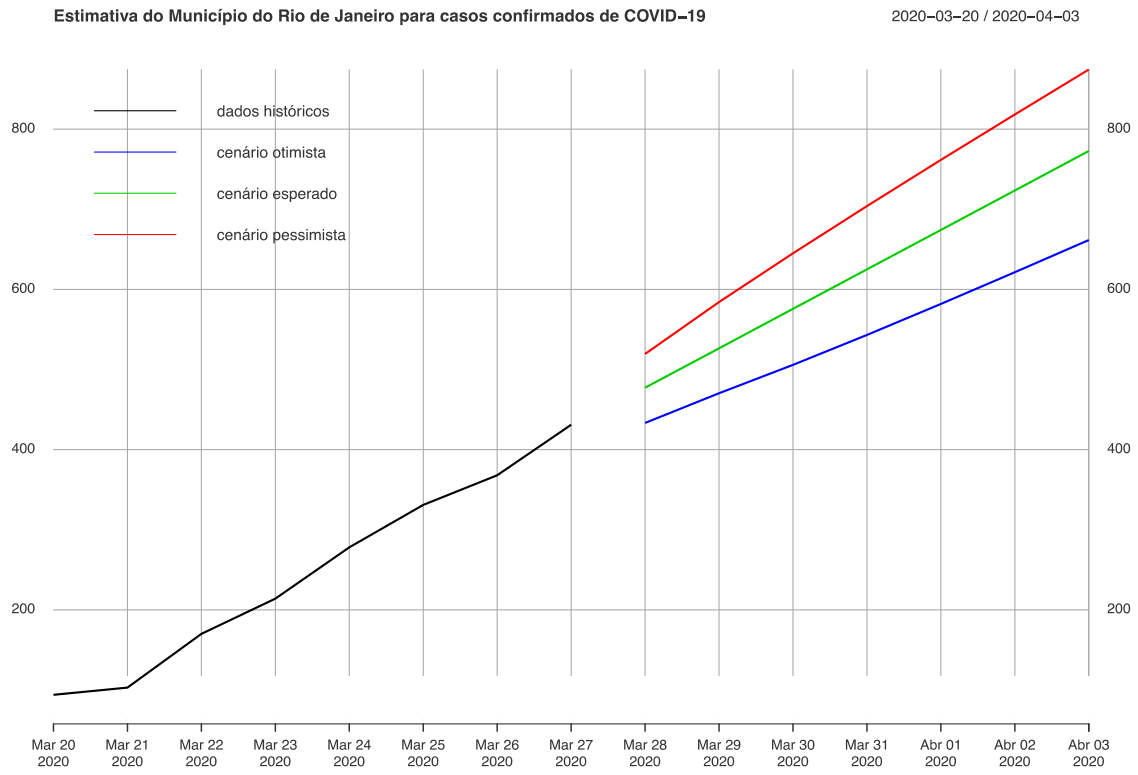
1) Teste de normalidade dos resíduos.
    a. ARIMA SO: p-valor =0,2037. Não podemos rejeitar a hipótese nula de normalidade.
    b. RL SO: p-valor =0,9506. Não podemos rejeitar a hipótese nula de normalidade.
    c. ARIMA SL: p-valor = 0,07444. Não podemos rejeitar a hipótese nula de normalidade.

d. RL SL: p-valor = 0,3007. Não podemos rejeitar a hipótese nula de normalidade.

2) Teste de autocorreção dos resíduos (ACF). Nenhum dos resíduos das séries supracitadas apresentaram auto correlação.
3) Menor média percentual absoluta do erro (MAPE) e menor erro quadrático médio (RMSE) dentre os modelos aprovados pelos critérios anteriores.

MODELOS	ME	RMSE	MAE	MPE	MAPE
ETS SO	-42,13	48,41	42,13	-22,22	22,22
ARIMA SO	-0,01	17,77	13,61	1,09	**7,0**
RL SO	0	**11,73**	9,69	-1,4	7,35
ETS SL	-42,13	48,42	42,13	-22,22	22,22
ARIMA SL	5,15	24,75	19,42	-0,82	8,92
RL SL	1,04	27,88	23,82	-0,56	9,7

Na tabela acima observa-se que o modelo ARIMA SO atende ao critério de menor MAPE, RL SO atende ao critério menor RMSE. Entretanto, com o objetivo de prover apenas uma previsão, combinamos os modelos ARIMA SO e RL SO considerando: o menor valor entre estas séries, a cada dia, para estimar a previsão otimista, o valor médio entre ambas para estimar o valor esperado por dia e o maior valor, a cada dia, entre elas para estimar a previsão pessimista. Assim, os próximos valores esperados de quantidade de casos confirmados de COVID-19 no Município do Rio de Janeiro em cada cenário são apresentados na figura e tabela abaixo.



DIAS	CENÁRIO OTIMISTA	CENÁRIO ESPERADO	CENÁRIO PESSIMISTA
28/03/2020	433	477	519
29/03/2020	470	526	584
30/03/2020	506	576	645
31/03/2020	543	625	704
01/04/2020	582	674	762
02/04/2020	621	723	818
03/04/2020	662	773	874

O objetivo dos modelos utilizados foi estimar com a maior precisão possível o total de casos de pacientes com COVID-19 em um curto horizonte de tempo e, com base no novo critério de corte dos dados históricos e novas observações, concluímos que a combinação dos modelos lineares se mostrou mais adequada.
Não é possível, no entanto, extrapolar este método para um período maior de tempo e nem mesmo concluir, a partir de seus resultados, em que estágio o município está na curva esperada de uma epidemia. Estudos com estes objetivos necessitam de abordagens mais complexas e também são conduzidas dentro do LEGOS.

Autor:
Daniel Assad
Currículo Lattes: http://lattes.cnpq.br/0258423859812498

Coordenação do LEGOS|UERJ
Profa Thaís Spiegel, DSc. | thais@eng.uerj.br
Currículo Lattes: http://lattes.cnpq.br/8880192361495671

Nota técnica divulgada em 28 de março de 2020.