Mapeamento e análise espacial, do número de casos e óbitos de COVID-19 no estado do Rio de Janeiro

Período de março a dezembro de 2020

FGV
Rapaz Nunes - **TG - 1400000**

INTRODUÇÃO

Foi registrado em 26 de janeiro de 2020, em São Paulo, o primeiro caso de COVID-19 no Brasil e na América do Sul. Um homem de 61 anos que havia retornado da cidade de Wuhan, na China, onde o novo coronavírus foi identificado pela primeira vez, apresentou sintomas, mas não foi a ser de 2020, confirmado por RT-PCR para o vírus SARS-CoV-2 (Bastos et al. 2020). Em 11 de março, a OMS declarou a COVID-19 como uma pandemia, quando o número de casos fora de territórios chegava em 118 casos e 4291 mortes.

Em 16 de março, foi registrado o primeiro caso de óbito no país, de um homem, e o número subiu em cinco vezes em dois dias de atraso, não o suficiente para conter o número de casos. Em um período de 30 dias, o número de casos subiu para 19.638 casos e 1.057 mortes. Quando apenas alguns dos estados do Rio Paulo com aproximadamente 45 mil mortes. Portal dessa pandemia no Brasil ultrapassou a marca dos 100 mil mortes.

A cidade do Rio de Janeiro, que está na região econômica e espacial de seu entorno(Fonte com outros estados, como Minas Gerais e São Paulo, e a cidade de Rio de Janeiro, que é a segunda cidade mais populosa do país, e a segunda maior metrópole, tem importância de se compreender o período de transmissão da doença, para melhor implementação de medidas de controle. Relevância e escala como pessoas que se deslocam entre cidades e estados.

OBJETIVO E METODO

Investigamos, com o uso de técnicas de análise exploratória e geoestatística espacial nos dados demográficos, epidemiológicos e de mobilidade, a evolução da COVID-19, no período de março de 2020 até o dia 31 de dezembro, quando o número de COVID-19 atinge a marca de 100 mil.

Foram utilizados os seguintes dados, disponibilizados pelo Google:

- **mob_state**: Índices de saída Relacionados com COVID-19 por estado (**mobilidade entre estados**)
- **mob_state**: Índices de chegada Relacionados com COVID-19 por estado (**que outros estados são percebidos**)
- **mobility**: Índice de mobilidade de acordo com 10 mobilidades
- **mapa_brasil**: Mapa do Brasil com as suas UF e mesorregiões
- **mapa_rio**: Mapa do Rio de Janeiro com as suas mesorregiões, 1 mesorregião, 1 mesorregião
- **dados2020**: Dados de desenvolvimento humano para cada UF(12 de janeiro de 2020) (**Fonte: Wikipedia**)
- **pop_municipios**: Número da população de interior
- **pop_uf**: Possuindo da população com acima a cada de estação baixa
- **pop_rmrj**: Possuindo da população com repetido no estado (**Fonte: Wikipedia**)
- **pop_rio**: Possuindo da população com repetido no estado (**Fonte: Wikipedia**)
- **pop2019**: Estimativa da população no IBGE para cada UF(**Fonte: Wikipedia**)

Para isso analisa, está ainda apresentando gráficos, mapas formais e modelos, e seus resultados serão comentados, no período de março de 2020 até o dia 31 de dezembro, quando o número de casos e mortes atinge a marca de 100 mil. Além disso, os dados de óbito e casos, com seus índices, bem como, índice de mobilidade e de saída e chegada de pessoas, serão comparados com os dados de outros estados do Brasil.

AVALIAÇÃO VISUAL DO NUMERO TOTAL DE CASOS - MARÇO A DEZEMBRO DE 2020.

No mapa abaixo, podemos observar o norte e o nordeste e grande e o centro-leste de casos por municípios.

Sabemos que **maiores números de casos**, se aprimorar todo de maneira e desperdício, foram registrados nas mesorregiões do Rio de Janeiro, Baixadas, Grande Rio e Grande, Belém Pará, Belém Pará Belém Pará, Belém Pará, Belém Pará,

As correlações entre as features se mantem nos três períodos, com destaque para as variáveis: **total_obitos, leitos_sus, resps_total, pop_rdonaca e prop_cad** que apresentam autocorrelação muito considerável.

MAPAS DE VIZINHANÇA

Foram criadas matrizes de vizinhanças para possibilitar análise das relações espaciais entre os casos de COVID-19, e fatores utilizados os critérios de contiguidade e de proximidade.

No critério de contiguidade são classificadas as fronteiras terrestres e no **de proximidade** a distância entre os centroides dos polígonos que determinam as municípios.

O município de Niterói e o São Gonçalo, por não possuírem fronteiras terrestres no caso de contiguidade e por distância, motivados pelo centroide do RJ que demanda se localizar muito distante do centro da cidade, foram tratados diferente sobres para a o (número de vizinhos por distância). Observamos que Niterói e Rio de Janeiro se tornam vizinhos a partir de k = 11.

Essa variação pode impactar consideravelmente os resultado dos modelos e do teste de Moran. Quanto maior o valor de K (para são do mesoestado para AC), perdem mais baixo o valor de p. Para este estudo escolhemos o valor intermediário k = 7.

Contiguidade (Queen)

Queen

Proximidade (knn2nb)

Proximidade

TESTES DE MORAN E MODELAGEM PARA TOTAL DE CASOS E TOTAL DE ÓBITOS

Utilizamos os testes de MORAN para avaliar as relações espaciais entre as variáveis, com a aplicação desses testes comprovamos a existência de autocorrelação espacial entre as variáveis de interesse.

Notamos que em quase todos os casos podemos rejeitar a hipótese nula de não existência de autocorrelação espacial, de forma que se rejeita um formação de clusters e que não nos permite considerada-las independentes. Em todos os testes para óbitos rejeitamos H0, no entanto nos testes de casos o p-value fica acima de 0,05. Nesse caso se recomenda o uso de modelos espaciais. Não rodamos diferença considerável para os critérios de vizinhança por proximidade não houve diferença entre utilizar o parâmetro k= 7.

TOTAL DE CASOS

```
##
##  Moran I test under randomisation
##
## data:  h2ohpstotal_casos
## weights: W.Queen.peso9
##
## Moran I statistic standard deviate = 1.4325, p-value = 0.076
## alternative hypothesis: greater
## sample estimates:
##Moran I statistic       Expectation          Variance
##     0.022605859      -0.010989011      0.000589946
```

```
##
##  Moran I test under randomisation
##
## data:  h2ohpstotal_casos
## weights: W.Queen.peso9
##
## Moran I statistic standard deviate = 1.5997, p-value = 0.05463
## alternative hypothesis: greater
## sample estimates:
##Moran I statistic       Expectation          Variance
##     0.025735108      -0.010989011      0.000556695
```

```
summary(ajuste_casosCAR1)

##
##  Call:
##  spautolm(formula = log(total_casos) ~ inv + mpt(prop_dbranca) +
##     prop_cad, data = base_covid_2, listw = W.Queen.peso9, family = "CAR")
##
##  Residuals:
##     Min       1Q   Median       3Q      Max
##  -0.2969778 -0.325091  0.030291  0.470550  1.652182
##
##  Coefficients:
##              Estimate Std. Error z value  Pr(>|z|)
##  (Intercept)  7.4467638  0.4914163 15.1476 < 2.2e-16
##  inv          1.1199366  1.0924188  1.0252  0.3053718
##  mpt(prop_dbranca) 0.0040120  0.0006723  5.9686 < 2.2e-16
##  prop_cad     -1.3844278  1.1373193 -1.2173  0.2235951
##
##  Lambda: 0.051595 LR test value: 0.1174 p-value: 0.73107
##  Numerical Hessian standard error of lambda: 0.242851
##
##  Log likelihood: -112.3366
##  ML residual variance (sigma squared): 0.64996, (sigma: 0.81661)
##  Number of observations: 92
##  Number of parameters estimated: 6
##  AIC: 236.27
```

```
summary(ajuste_casosCAR2)

##
##  Call:
##  spautolm(formula = log(total_casos) ~ inv + mpt(prop_dbranca) +
##     prop_cad, data = base_covid_2, listw = W.Queen.peso9, family = "CAR")
##
##  Residuals:
##     Min       1Q   Median       3Q      Max
##  -0.2970128 -0.324997  0.032028  0.469750  1.646097
##
##  Coefficients:
##              Estimate Std. Error z value  Pr(>|z|)
##  (Intercept)  7.4467638  0.4914163 15.1476 < 2.2e-16
##  inv          1.1199366  1.0927945  0.9979  0.3184251
##  mpt(prop_dbranca) 0.0040120  0.0006723  5.9686 < 2.2e-16
##  prop_cad     -1.3844278  1.1354160 -1.2193  0.2228652
##
##  Lambda: 0.021106 LR test value: 0.12576 p-value: 0.72286
##  Numerical Hessian standard error of lambda: 0.262841
##
##  Log likelihood: -112.3324
##  ML residual variance (sigma squared): 0.64930, (sigma: 0.81589)
##  Number of observations: 92
##  Number of parameters estimated: 6
##  AIC: 236.26
```

```
summary(ajuste_casosCAR3)

##
##  Call:
##  spautolm(formula = log(total_casos) ~ mpt(inv) + mpt(prop_dbranca) +
##     mpt(prop_cad), data = base_covid_2, listw = W.Queen.peso9,
##     family = "CAR")
##
##  Residuals:
##     Min       1Q   Median       3Q      Max
##  -2.5172368 -0.251124  0.019601  0.445689  1.587219
##
##  Coefficients:
##              Estimate Std. Error z value  Pr(>|z|)
##  (Intercept)  6.5501885  0.3473625  8.8388 < 2.2e-16
##  mpt(inv)     1.5301493  2.0124348  0.7604  0.4469927
##  mpt(prop_dbranca) 0.0040120  0.0006723  5.9686 < 2.2e-16
##  mpt(prop_cad) -4.6776149  1.0354603 -4.5156 6.3955868
##
##  Lambda: 0.020994 LR test value: 0.21857 p-value: 0.64013
##  Numerical Hessian standard error of lambda: 0.262321
##
##  Log likelihood: -111.1461
##  ML residual variance (sigma squared): 0.6572, (sigma: 0.81074)
##  Number of observations: 92
##  Number of parameters estimated: 6
##  AIC: 234.3
```

```
AIC(ajuste_casosCAR1, ajuste_casosCAR2, ajuste_casosCAR3, ajuste_casosCAR3)

##                df      AIC
## ajuste_casosCAR1  6 236.2732
## ajuste_casosCAR2  6 236.2648
## ajuste_casosCAR3  6 234.2922
## ajuste_casosCAR3  6 234.2922
```

TOTAL DE ÓBITOS

```
##
##  Moran I test under randomisation
##
## data:  h2ohpstotal_obitos
## weights: W.Queen.peso9
##
## Moran I statistic standard deviate = 2.4599, p-value = 0.0002702
## alternative hypothesis: greater
## sample estimates:
##Moran I statistic       Expectation          Variance
##     0.156425656      -0.010989011      0.001086449
```

```
##
##  Moran I test under randomisation
##
## data:  h2ohpstotal_obitos
## weights: W.Queen.peso9
##
## Moran I statistic standard deviate = 2.7862, p-value = 0.0005288
## alternative hypothesis: greater
## sample estimates:
##Moran I statistic       Expectation          Variance
##     0.161116297      -0.010989011      0.000592148
```

```
summary(ajuste_obitosCAR1)

##
##  Call:
##  spautolm(formula = log(total_obitos) ~ mpt(inv) + mpt(prop_dbranca) +
##     mpt(prop_cad), data = base_covid_2, listw = W.Queen.peso9,
##     family = "CAR")
##
##  Residuals:
##     Min       1Q   Median       3Q      Max
##  -0.6173268 -0.433377  0.079548  0.590161  1.996416
##
##  Coefficients:
##              Estimate Std. Error z value  Pr(>|z|)
##  (Intercept)  5.3412445  0.4114664 12.9806 < 2.2e-16
##  mpt(inv)     -0.0290536  0.0142502 -2.0389 0.0414371
##  mpt(prop_total) 0.0042973  0.0139263  0.2753 0.7946-10
##  mpt(prop_dbranca) 0.0024177  0.0004842  4.9934e-16
##  mpt(prop_cad) -4.6460425  1.1919562 -3.7733 0.0001633
##
##  Lambda: 0.29453 LR test value: 4.016 p-value: 0.04517
##  Numerical Hessian standard error of lambda: 0.14392
##
##  Log likelihood: -108.6294
##  ML residual variance (sigma squared): 0.599, (sigma: 0.77395)
##  Number of observations: 92
##  Number of parameters estimated: 7
##  AIC: 231.27
```

```
summary(ajuste_obitosCAR2)

##
##  Call:
##  spautolm(formula = log(total_obitos) ~ mpt(inv) + mpt(prop_total) +
##     mpt(prop_dbranca) + mpt(prop_cad), data = base_covid_2,
##     listw = W.Queen.peso9, family = "CAR")
##
##  Residuals:
##     Min       1Q   Median       3Q      Max
##  -0.6173268 -0.433378  0.082635  0.592152  1.422667
##
##  Coefficients:
##              Estimate Std. Error z value  Pr(>|z|)
##  (Intercept)  5.3412445  0.4114664 12.9806 < 2.2e-16
##  mpt(inv)     -0.0290536  0.0142502 -2.0389 0.0414371
##  mpt(prop_total) 0.0042973  0.0139263  0.2753 2.9946-10
##  mpt(prop_dbranca) 0.0024177  0.0004842  4.9934e-16
##  mpt(prop_cad) -4.6460425  1.1919562 -3.7733 0.0001633
##
##  Lambda: 0.27997 LR test value: 0.7276 p-value: 0.3936934
##  Numerical Hessian standard error of lambda: 0.227029
##
##  Log likelihood: -108.2996
##  ML residual variance (sigma squared): 0.59391, (sigma: 0.77066)
##  Number of observations: 92
##  Number of parameters estimated: 7
##  AIC: 230.59
```

```
summary(ajuste_obitosCAR3)

##
##  Call:
##  spautolm(formula = log(total_obitos) ~ mpt(inv_max) + mpt(prop_total) +
##     mpt(prop_dbranca) + mpt(prop_cad), data = base_covid_2,
##     listw = W.Queen.peso9, family = "CAR")
##
##  Residuals:
##     Min       1Q   Median       3Q      Max
##  -1.68893 -0.43375  0.14042  0.62395  1.43679
##
##  Coefficients:
##              Estimate Std. Error z value  Pr(>|z|)
##  (Intercept)  5.1987575  0.4810656  4.8763 5.3896-11
##  mpt(inv_max) -0.0300724  0.0141228 -2.1293 0.0332255
##  mpt(prop_total) 0.0074327  0.0139264  0.5336 0.5936-10
##  mpt(prop_dbranca) 0.0024187  0.0004842  5.2041 2.1279e-13
##  mpt(prop_cad) -4.4940425  1.1919562 -3.7708 0.0001633
##
##  Lambda: 0.13896 LR test value: 7.4767 p-value: 0.0055939
##  Numerical Hessian standard error of lambda: 0.226012
##
##  Log likelihood: -107.815
##  ML residual variance (sigma squared): 0.56702, (sigma: 0.75301)
##  Number of observations: 92
##  Number of parameters estimated: 7
##  AIC: 229.63
```

```
summary(ajuste_obitosCAR3)

##
##  Call:
##  spautolm(formula = log(total_obitos) ~ mpt(inv_max) + mpt(prop_total) +
##     mpt(prop_dbranca) + mpt(prop_cad), data = base_covid_2,
##     listw = W.Queen.peso9, family = "CAR")
##
##  Residuals:
##     Min       1Q   Median       3Q      Max
##  -1.68893 -0.43375  0.14042  0.62395  1.43679
##
##  Coefficients:
##              Estimate Std. Error z value  Pr(>|z|)
##  (Intercept)  5.1987575  0.4810656  4.8763 5.3896-11
##  mpt(inv_max) -0.0300724  0.0141228 -2.1293 0.0332255
##  mpt(prop_total) 0.0074327  0.0139264  0.5336 0.5936-10
##  mpt(prop_dbranca) 0.0024187  0.0004842  5.2041 2.1279e-13
##  mpt(prop_cad) -4.4940425  1.1919562 -3.7708 0.0001633
##
##  Lambda: 0.13896 LR test value: 7.4767 p-value: 0.0055939
##  Numerical Hessian standard error of lambda: 0.226012
##
##  Log likelihood: -107.815
##  ML residual variance (sigma squared): 0.56702, (sigma: 0.75301)
##  Number of observations: 92
##  Number of parameters estimated: 7
##  AIC: 229.63
```

```
AIC(ajuste_obitosCAR1, ajuste_obitosCAR2, ajuste_obitosCAR3, ajuste_obitosCAR3)

##                df      AIC
## ajuste_obitosCAR1  7 231.2732
## ajuste_obitosCAR2  7 230.5918
## ajuste_obitosCAR3  7 229.6305
## ajuste_obitosCAR3  7 229.6305
```

ANÁLISE DOS MODELOS

Foram avaliados os seguintes dois modelos de regressão linear multipla, SAR e CAR com matrizes de vizinhança W e B. Onde W pesa pesos divididos igualmente entre as vizinhos de cada região e B considerando peso 1 independente da quantidade de vizinhos que município possua.

- Com todos os ajustes o modelo **ajuste_casosCAR2** apresentou o menor AIC: 234.0866 no modelagem para total de casos.
- Já na modelagem para total de óbitos dois modelos apresentaram AIC's quase iguais: 229.6301 e ambos foram os menores números encontrados com todos os ajustes, sendo eles: **ajuste_obitosCAR1 e ajuste_obitosCAR1**.

CONCLUSÃO

Concluímos que os dados apresentados estão agregados por área e por este motivo os modelos CAR e SAR foram utilizados. Comparamos índices e testamos as correlações com os totais de casos e totais de óbitos para a seleção de modelos que melhor se ajustassem. Foi possível concluir que há correlação espacial entre os municípios e não seria razoável considerar que as unidades fossem independentes e portanto não seria recomendado utilizar modelos de regressão linear.

Um ponto próprio futuro seria interessante utilizar um modelo de séries temporais para estimar e prever a quantidade de pessoas infectadas e óbitos nos primeiros e segundos ciclos meses de 2021.