

Universidade de São Paulo - USP
Instituto de Ciências Matemáticas e de Computação - ICMC
Departamento de Matemática Aplicada e Estatística - SME
Introdução à Inferência Estatística

Trabalho 1

Gabriel Passarelli - Número USP: 11218480
Hugo Gili Zancul - Número USP: 10295815
Karina Drews Bernardi Ferreira - Número USP: 10691859
Stéfane Tame Monteiro Oliveira - Número USP: 10829970
Profa. Dra. Katiane Silva Conceição

Maio
2021

Conteúdo

1	Classificação da variável	1
2	Visualização gráfica da série de dados	2
3	Estatísticas das variáveis	6
3.1	Média, mediana e quantis	6
3.2	Amplitude total, amplitude interquartílica e desvio-padrão . .	8
4	Visualização gráfica da série de dados diários - Boxplots e Histogramas	10
5	Tabelas de frequência	13
6	Código	17

1 Classificação da variável

Este trabalho buscou analisar a pandemia de COVID-19 na região Sudeste do Brasil no período entre 25/02/2020 e 29/04/2021. Essa avaliação foi feita com base em duas variáveis: número de óbitos por dia e número de casos diários registrados em cada estado da região. Para que fosse possível uma melhor comparação entre os locais analisados, os dados foram convertidos em um equivalente proporcional para cada 100.000 habitantes utilizando a seguinte fórmula:

$$\frac{\text{Dados} \times 100.000}{\text{População do estado}}.$$

Dessa forma, o que antes eram duas variáveis quantitativas discretas, por se tratarem de um número de pessoas - casos positivos da doença e óbitos em decorrência dela -, se tornaram, após a conversão dos dados, duas variáveis quantitativas contínuas (são aquelas que tratam de características mensuráveis que assumem valores na reta real).

2 Visualização gráfica da série de dados

A visualização gráfica dos dados de cada estado pode ser feita com o auxílio da função *plot()*. A obtenção dos gráficos que seguem é bem direta: basta passar o vetor com os registros de casos (ou de óbitos) como argumento da função, e também as legendas de cada eixo. Faremos uma análise mais detalhada do processo de obtenção dos gráficos para as curvas das variáveis acumuladas.

Aqui cabe dizer que seguiremos um padrão durante o restante do texto: o Rio de Janeiro é representado pela cor verde nos gráficos; Minas, azul; São Paulo, vermelha; e Espírito Santo, roxa.

Os gráficos que mostram o dia a dia da pandemia em cada estado são difíceis de serem analisados através da observação direta, devido à alta variabilidade dos registros, principalmente no estado do Rio de Janeiro. Contudo, ainda pode-se observar, em especial, nos gráficos do Espírito Santo, algumas regiões onde os pontos elevam-se, indicando taxas de contágio e de óbito mais acentuadas, sendo possível inclusive observar alta correlação entre as duas variáveis.

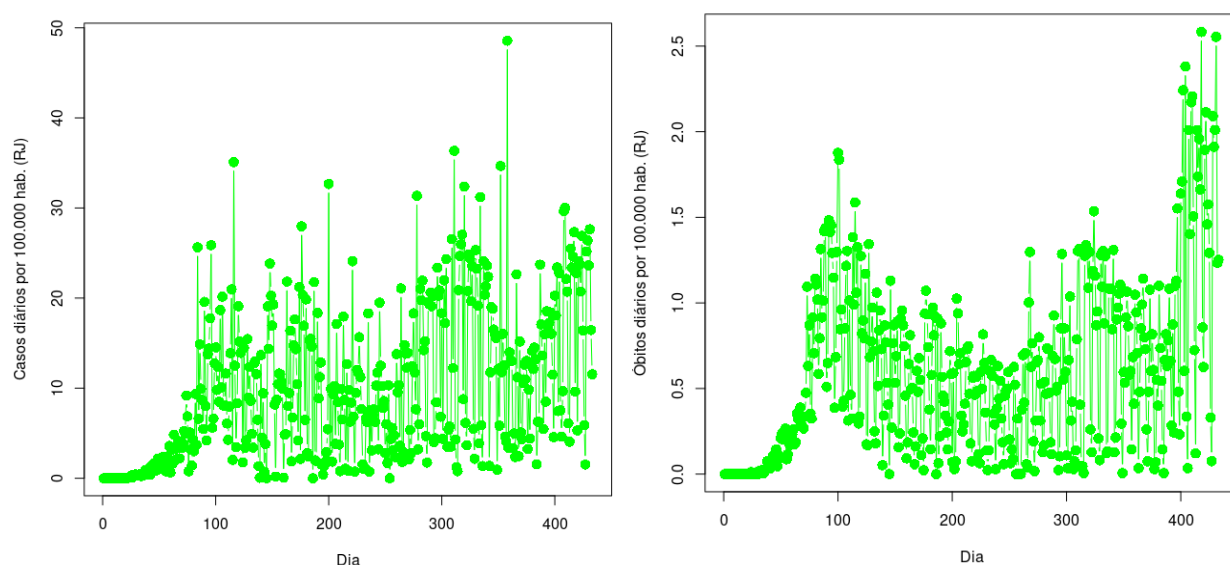


Figura 1: Rio de Janeiro

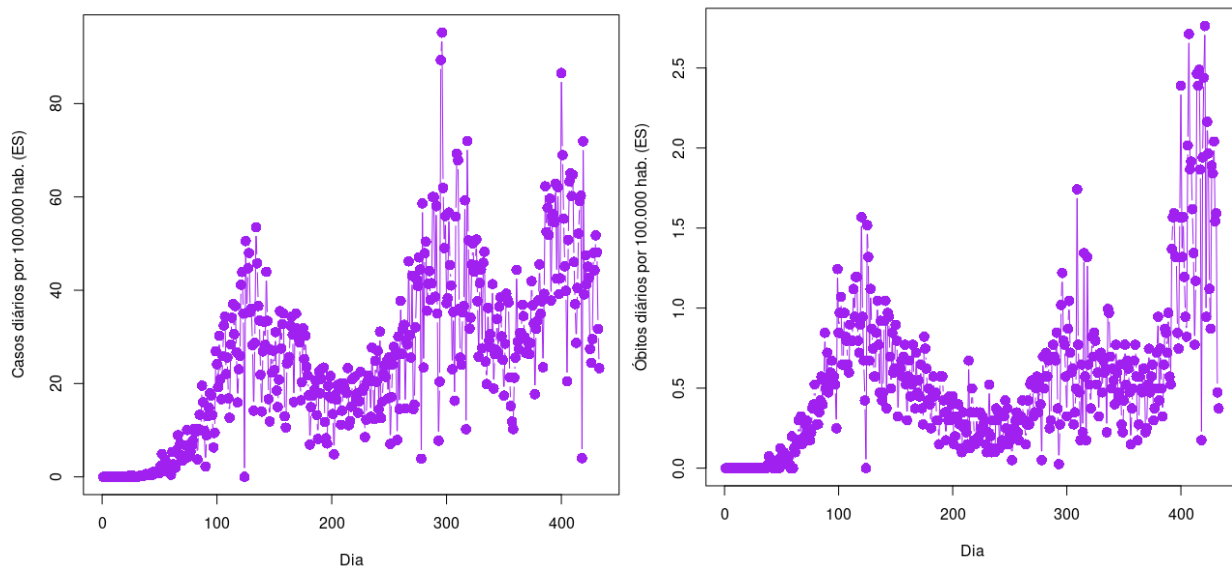


Figura 2: Espírito Santo

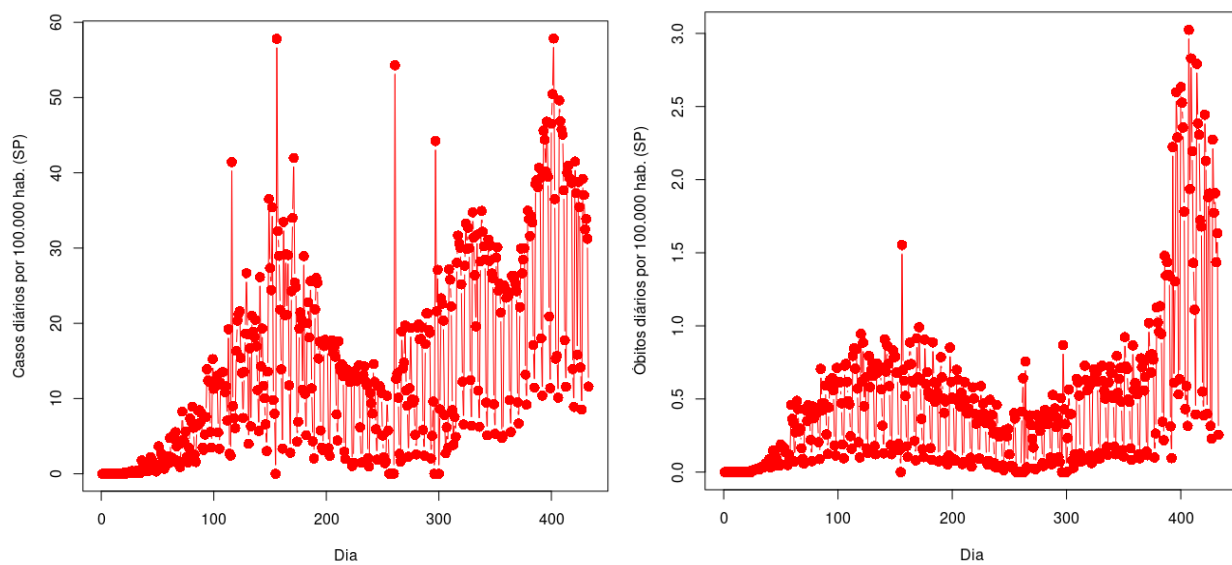


Figura 3: São Paulo

Passemos agora aos gráficos das variáveis acumuladas. Antes, porém, iremos explicar brevemente como o trecho do código que gera esses gráficos, anexado ao relatório, funciona (ele está compreendido exatamente entre os comentários 'Gráfico Acumulado' e 'Gráfico casos diários/obitos diários').

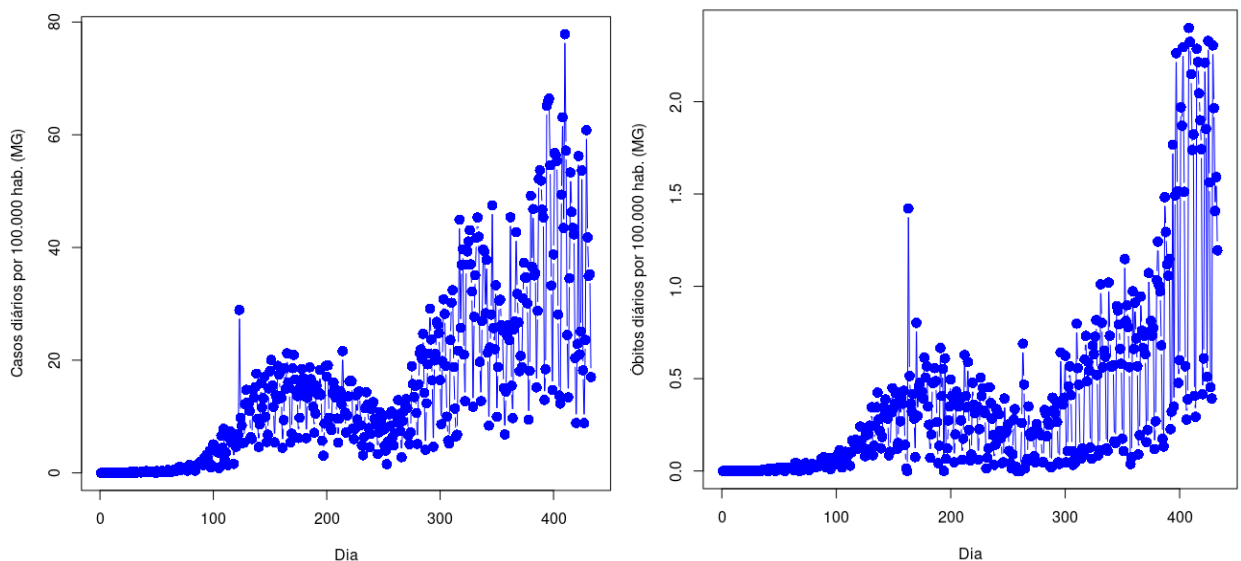
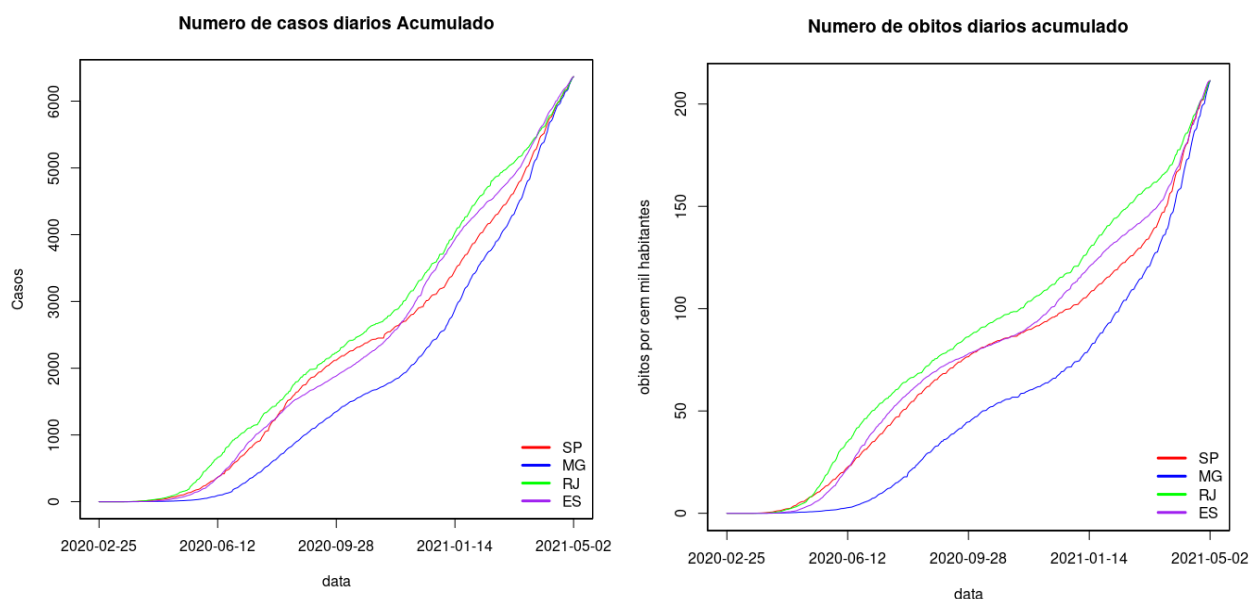


Figura 4: Minas Gerais

Para não precisarmos gerar um gráfico para a variável de cada estado, executamos a linha `par(new = TRUE)`, que define como *TRUE* o parâmetro *new*, e isso faz com que o próximo gráfico gerado pela próxima ocorrência da função `plot()` no código sobreponha o gráfico atual. Por esse motivo, inclusive, depois da linha 14, a cada chamada da função `plot()`, passamos “*n*” como valor dos parâmetros *xaxt* e *yaxt*, de modo que não se visualize os eixos do novo gráfico (lembre-se de que, como estamos sobrepondo os gráficos, só precisamos gerar os eixos uma única vez); além desses dois parâmetros, temos também *xlab* e *ylab* (os quais também recebem ‘*n*’ depois da linha 14), que servem para se definir as legendas de cada eixo do gráfico, e *type*, que define o estilo de visualização dos dados (“*l*” indica que os valores de nosso vetor serão ligados por retas, e assim obtemos um gráfico “contínuo”, como se vê na figura 2).

Finalmente, `axis()` na linha 10 gera o eixo x do gráfico, enquanto `legend()` gera a legenda. A função `x11()`, na linha 2, serve apenas para que uma janela de visualização seja criada e assim possamos visualizar os gráficos mais facilmente.

Analisando cada gráfico, podemos localizar no tempo as ‘ondas’ da pandemia: os períodos de aumento e de diminuição das taxas de contágio e de óbitos. De fato, se olharmos para Minas Gerais a partir do final de junho até o começo de outubro, vê-se uma curva inclinada de casos e de óbitos, que, após esse período, entra em uma fase de crescimento mais lento, pois torna-se mais suave, indicando assim o fim da ‘primeira onda’. Contudo, tal período



é curto em relação à duração total da série temporal. Em janeiro de 2021 já tem-se novamente uma trajetória acendente de casos e óbitos, que não apresentava indícios de arrefecimento até o final da série. Essa é, portanto, a segunda onda, que tem duração mais prolongada do que a primeira, e com períodos de taxas de contágio e de óbitos mais altas do que as da primeira onda, como se observa pela inclinação das curvas.

Podemos estender essa análise para as curvas de São Paulo e para as do Espírito Santo. As do Rio de Janeiro, por outro lado, evidenciam que a pandemia nesse estado atingiu sua população de maneira mais intensa em 2020 do que nos demais estados da região Sudeste, sem períodos claros de diminuição das taxas de contágio e de mortes. Entretanto, no final da série, o padrão de comportamento das curvas se inverte, e há uma tendência de que a curva do Rio de Janeiro seja sobreposta pelas dos demais estados, evidenciando assim que o agravamento da pandemia em 2021 nos outros três estados compensa os números menores que eles mantinham até então. Isso traz nova confirmação, portanto, para o fato de que a segunda onda de casos e de óbitos na região sudeste, em especial em Minas Gerais, São Paulo e Espírito Santo, foi mais intensa do que a primeira. Por fim, a análise dos gráficos das variáveis acumuladas reforça a constatação feita no caso anterior de que as variáveis têm um nível de correlação muito alto (fato intuitivo e que, evidentemente, já era esperado).

3 Estatísticas das variáveis

Todas as estatísticas foram calculadas com auxílio das funções nativas do R, como se vê pelo código anexado.

3.1 Média, mediana e quantis

Vamos agora fazer uma análise comparativa das variáveis de casos e de mortes em cada estado através do estudo dos valores de suas respectivas médias, medianas e quantis. Lembramos que as variáveis (e, portanto, as tabelas que seguem) medem o número de óbitos ou de casos sempre por centena de milhar de habitantes do seu respectivo estado.

Na tabela 1, vemos que a variável de óbitos teve um comportamento bem próprio em cada estado. Vê-se, em particular, que no estado do Rio de Janeiro os registros são os mais altos da região. Isso se evidencia pela média, a mais alta dentre as demais, seguida pela do Espírito Santo, de São Paulo, e Minas Gerais, nessa ordem; e também pela mediana e pelo terceiro quartil: 75% dos registros nesse estado são inferiores a 0.9 óbitos diários por centena de milhar de habitante; no Espírito Santo, o valor cai para 0.722; em São Paulo, 0.638; e em Minas Gerais, temos 0.477, o que reforça que em Minas tivemos a menor mortalidade da doença da Região.

	MG	SP	ES	RJ
Mínimo	0	0	0	0
$q(25\%)$	0.038	0.100	0.224	0.145
Mediana	0.208	0.394	0.428	0.516
$q(75\%)$	0.477	0.638	0.722	0.904
Máximo	2.400	3.025	2.762	2.583
Média	0.374	0.488	0.549	0.600

Tabela 1: Comparação pelo número de óbitos diários

A tabela 2 nos mostra que o estado com maior número médio de casos, proporcionalmente à sua população, foi o Espírito Santo; em seguida, temos Minas Gerias e São Paulo, com valores variando em torno de 14.8 casos por centena de milhar de habitantes; e finalmente, o Rio de Janeiro, com a menor média de casos, 9.998. Comparando esses resultados com os da tabela anterior, vemos que, apesar da alta correlação, observada nos gráficos, do comportamento das variáveis de óbitos e casos, não se pode supor que a medida de casos forneça uma visão ampla da situação epidemiológica da doença dentro de um estado. Como evidencia não só a média de cada um dos estados, mas também o terceiro quartil, Minas Gerais possui registros mais altos

na variável de casos do que o Rio de Janeiro. Contudo, Minas Gerais contabilizou proporcionalmente ao Rio de Janeiro um número de óbitos inferior. O mesmo acontece quando comparamos as colunas de Minas Gerais em cada uma das tabelas com as de São Paulo.

Possíveis explicações para essa divergência entre as tabelas poderiam surgir casos investigássemos a maneira pela qual se deu, e com qual intensidade e amplitude, a realização de testes nos doentes em cada localidade. Em outras palavras, é natural supor que maiores números de casos devem vir acompanhados de maiores números de mortes. Contudo, se um estado realiza uma quantidade de testes insuficientes para capturar o real número de infectados entre sua população, a variável de casos evidentemente destoará daquilo que a de óbitos mostrará.

Além disso, podemos supor também que a distância entre as avaliações proporcionadas pela variável de casos em relação à de mortes seja afetada pelo fato de que a variável de casos não registra a letalidade da doença em cada região, ou seja, o número de óbitos pelo de infectados. Além da qualidade e da quantidade de medidas de controle da doença tomadas por cada estado durante a pandemia, fatores como acesso a saneamento, qualidade dos serviços públicos, como saúde e transporte, a porcentagem de trabalhadores com carteira, ou ainda, que podem trabalhar sem sair de casa, variam imensamente a depender da localidade, e também impactam diretamente como a doença se propaga entre os habitantes dessa localidade.

	MG	SP	ES	RJ
Mínimo	0	0	0	0
$q(25\%)$	3.345	3.480	11.845	2.786
Mediana	11.451	11.720	24.060	8.120
$q(75\%)$	20.785	22.790	36.604	15.668
Máximo	77.846	57.860	95.210	48.567
Média	14.949	14.700	25.210	9.998

Tabela 2: Comparação pelo número de casos diários

3.2 Amplitude total, amplitude interquartílica e desvio-padrão

Na tabela 3, temos a amplitude total, amplitude interquartílica e desvio-padrão para os números de óbitos diários registrados em cada estado. A amplitude total é similar em todos os estados observados - por volta de 2.5 -, mas em São Paulo esse valor destoa e é um pouco maior com relação aos outros. Essa amplitude maior, enquanto dado isolado, não significa muito - poderia ter sido gerada por um *outlier* ou ser evidência de uma disparidade concreta no número de óbitos em dias distintos durante o período em que os dados foram coletados.

A amplitude interquartílica (que representa a amplitude da faixa onde estão localizados 50% dos valores coletados mais próximos à mediana) é similar tanto em MG, quanto em SP e no ES: bem próxima de 0.5, o que, juntamente com o valor da amplitude total, indica que há uma concentração dos valores coletados em torno da mediana. Para o estado do RJ, a amplitude interquartílica foi de 0.759, evidenciando uma dispersão ligeiramente maior com relação aos outros locais, mas ainda indicando um acúmulo dos valores coletados em torno da mediana.

Quanto ao desvio-padrão, temos que é um valor relativamente alto em todos os lugares analisados - cerca de 0.5 em dados que possuem amplitude variando de 2.4 a 3.025 dependendo do estado-, ou seja, há uma dispersão grande dos dados coletados com relação à média. Este fato era esperado, uma vez que tivemos um início de pandemia com pouquíssimos registros de mortes e, conforme o país experienciou picos de contágio, foram registrados também picos de mortes. Além disso, contribui também o fato de que há um registro menor de mortes aos finais de semana (pela diminuição no número de funcionários em plantão nos hospitais e unidades de saúde), o que leva a uma disparidade em diferentes dias de uma mesma semana. Por exemplo, no período de uma semana, a terça-feira sempre terá um registro de óbitos maior do que o domingo, ainda que os dois dias pertençam a uma mesma fase da pandemia.

	MG	SP	ES	RJ
Amplitude total	2.400	3.025	2.762	2.583
Amplitude inter.	0.439	0,538	0,498	0,759
Desvio-padrão	0.494	0.529	0.490	0.524

Tabela 3: Comparação pelo número de óbitos diários

Na tabela 4, temos a amplitude total, amplitude interquartílica e desvio-padrão para os números de casos diários registrados em cada estado. A

amplitude total varia muito, chegando a ser quase o dobro no ES com relação ao RJ, por exemplo. Possíveis justificativas são: uma maior disseminação da COVID-19 em determinados estados, uma testagem ampla e organizada em certos locais, uma combinação destes dois fatores, entre outros pontos.

A amplitude interquartílica segue uma tendência de ser cerca de 25% da amplitude total em cada estado, ou seja, indica que há uma concentração dos dados coletados em torno da mediana.

O desvio-padrão observado é alto, indicando uma dispersão dos dados coletados com relação à média. Esse fato pode ser explicado pela evolução da pandemia no país, o que elevou o número de casos diários por 100 mil habitantes de 0 a valores como 77.846 e 95.210 (em MG e ES, respectivamente). De modo semelhante ao número de óbitos diários, o número de casos também sofre distorções com relação aos poucos registros durante os finais de semana, uma vez que há menos profissionais trabalhando.

	MG	SP	ES	RJ
Amplitude total	77.846	57.860	95.210	48.567
Amplitude inter.	17.441	19.310	24.760	12.882
Desvio-padrão	14.940	12.876	18.092	8.595

Tabela 4: Comparação pelo número de casos diários

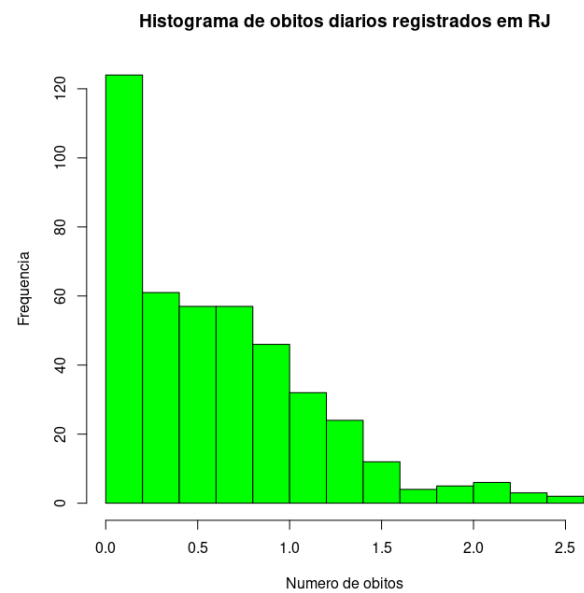
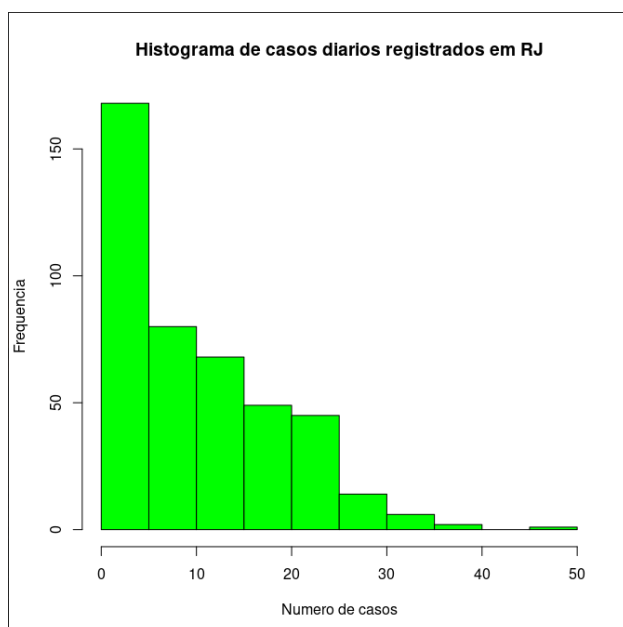
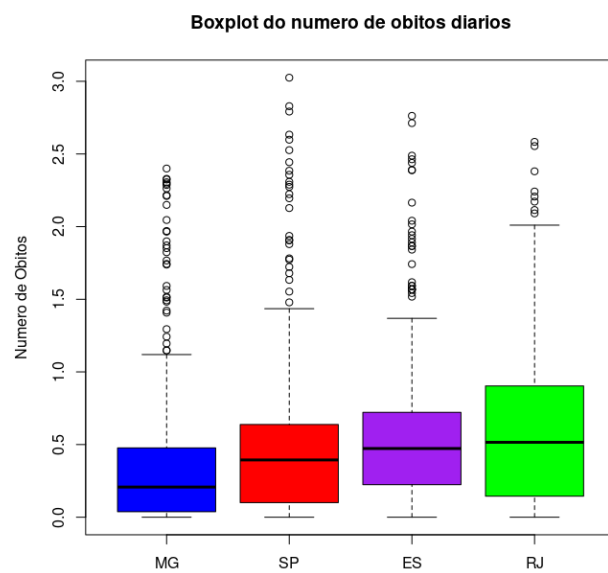
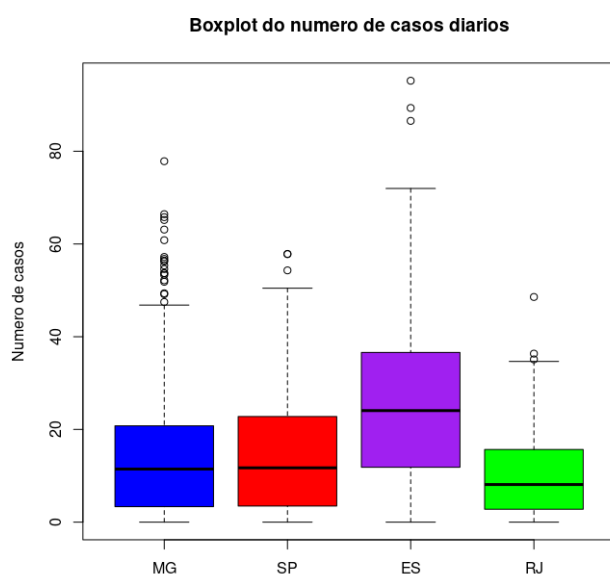
4 Visualização gráfica da série de dados diários - Boxplots e Histogramas

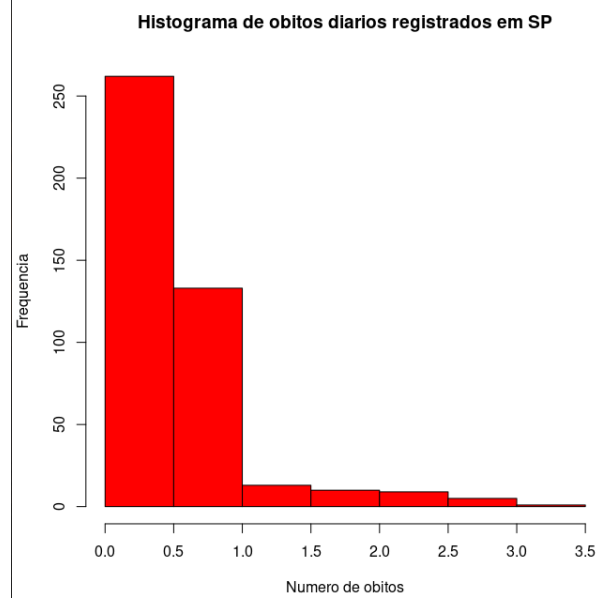
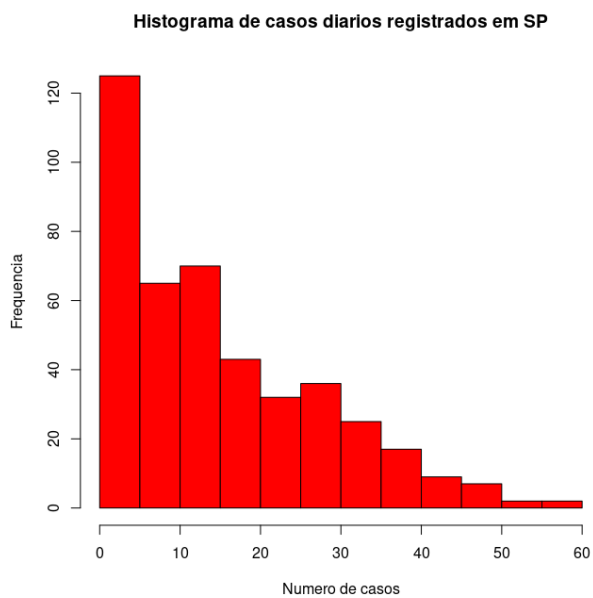
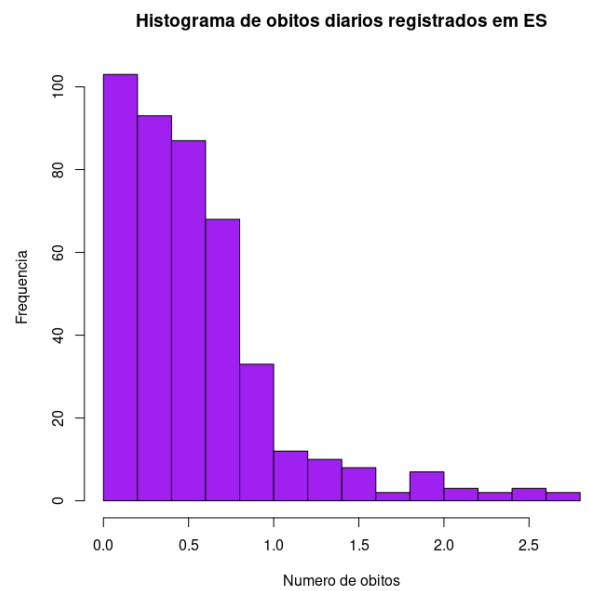
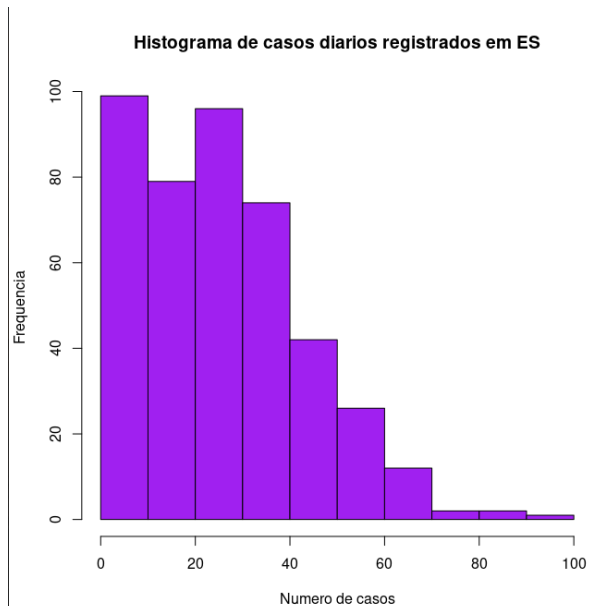
A avaliação dos gráficos em boxplot e em histograma nos permite concluir os mesmos fatos que observamos a partir das estatísticas sumarizadas em tabelas na seção anterior.

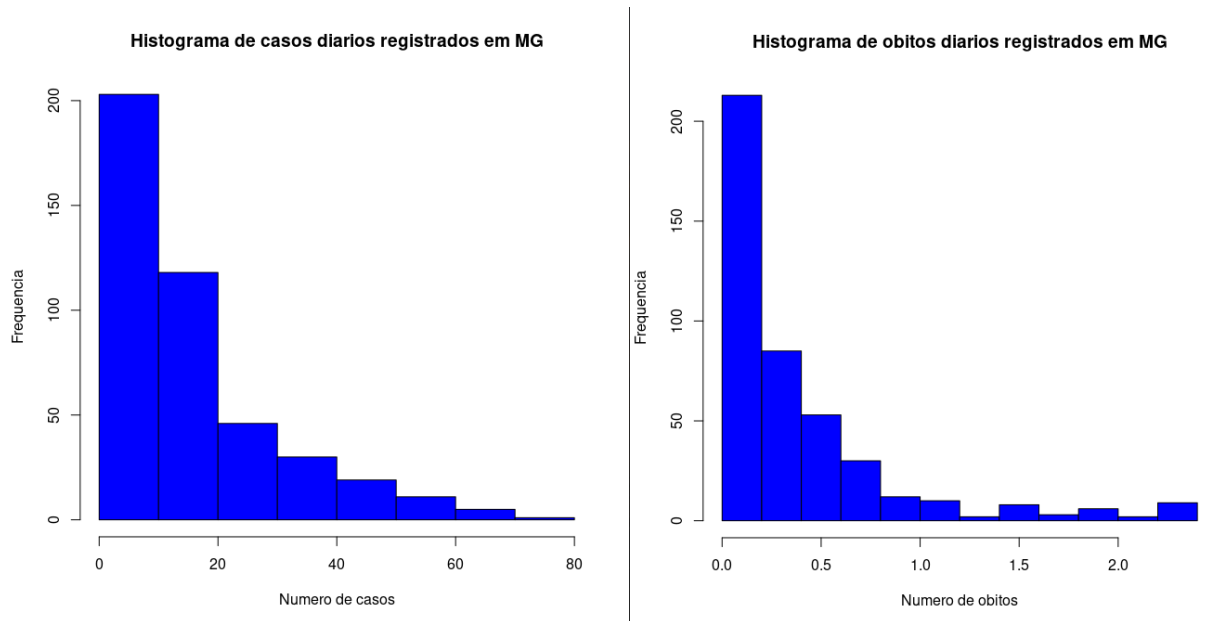
Nos boxplots do número de casos diários, dois fatores evidenciam a existência de pontos variados maiores que o terceiro quartil: é possível perceber a presença de vários *outliers*, especialmente em MG - todos acima do valor máximo -; as caudas das distribuições de todos os estados são compridas. Além disso, no RJ é possível perceber uma assimetria positiva, o que indica um grande número de dias com poucos casos.

Já nos boxplots do número de óbitos diários, é possível perceber uma enorme presença de *outliers* em todos os estados, todos acima do valor máximo. De forma análoga à análise anterior, as caudas acima do terceiro quartil são compridas, indicando a existência de registros diversos nessa região. O fato da mediana ser um valor tão baixo perto da amplitude máxima indica que a maioria absoluta de registros envolveu valores pequenos. Ou seja: os valores menores que a mediana se encontram em um pequeno intervalo, enquanto os valores maiores que a mediana estão espalhados em um grande intervalo - análise válida para os quatro estados.

Os histogramas nos dão uma visão mais clara de como seria a densidade de probabilidade das variáveis em cada estado. No geral, observa-se uma assimetria que desloca o gráfico para a esquerda, fazendo com que os valores se concentrem próximos ao zero - corroborando o que foi evidenciado nos boxplots.







5 Tabelas de frequência

Série de novos casos e óbitos no período 25/02/2020 - 29/04/2021.

Classes c_i	frequência absoluta	frequência relativa	frequência relativa (%)	frequência acumulada	frequência acumulada (%)
[0, 7.8624)	171	0.39	39.49	171	39.49
[7.8624, 15.725)	110	0.25	25.40	281	64.90
[15.725, 23.587)	59	0.14	13.63	340	78.52
[23.587, 31.45)	33	0.08	7.62	373	86.14
[31.45, 39.312)	22	0.05	5.08	395	91.22
[39.312, 47.175)	18	0.04	4.16	413	95.38
[47.175, 55.037)	9	0.02	2.08	422	97.46
[55.037, 62.899)	6	0.01	1.39	428	98.85
[62.899, 70.762)	4	0.01	0.92	432	99.77
[70.762, 78.624)	1	0.00	0.23	433	100.00

Tabela 5: Casos em Minas Gerais

Classes c_i	frequência absoluta	frequência relativa	frequência relativa (%)	frequência acumulada	frequência acumulada (%)
[0, 0.2424)	229	0.53	52.89	229	52.89
[0.2424, 0.4848)	97	0.22	22.40	326	75.29
[0.4848, 0.7271)	44	0.10	10.16	370	85.45
[0.7271, 0.9695)	21	0.05	4.85	391	90.30
[0.9695, 1.212)	12	0.03	2.77	403	93.07
[1.212, 1.454)	4	0.01	0.92	407	94.00
[1.454, 1.697)	6	0.01	1.39	413	95.38
[1.697, 1.939)	7	0.02	1.62	420	97.00
[1.939, 2.181)	4	0.01	0.92	424	97.92
[2.181, 2.424)	9	0.02	2.08	433	100.00

Tabela 6: Óbitos em Minas Gerais

Classes c_i	frequência absoluta	frequência relativa	frequência relativa (%)	frequência acumulada	frequência acumulada (%)
[0, 5.8435)	143	0.33	33.03	143	33.03
[5.8435, 11.687)	71	0.16	16.40	214	49.42
[11.687, 17.53)	63	0.15	14.55	277	63.97
[17.53, 23.374)	49	0.11	11.32	326	75.29
[23.374, 29.217)	41	0.09	9.47	367	84.76
[29.217, 35.061)	29	0.07	6.70	396	91.45
[35.061, 40.904)	20	0.05	4.62	416	96.07
[40.904, 46.748)	10	0.02	2.31	426	98.38
[46.748, 52.591)	4	0.01	0.92	430	99.31
[52.591, 58.435)	3	0.01	0.69	433	100.00

Tabela 7: Casos em São Paulo

Classes c_i	frequência absoluta	frequência relativa	frequência relativa (%)	frequência acumulada	frequência acumulada (%)
[0, 0.3055)	182	0.42	42.03	182	42.03
[0.3055, 0.611)	119	0.27	27.48	301	69.52
[0.611, 0.9165)	89	0.21	20.55	390	90.07
[0.9165, 1.222)	10	0.02	2.31	400	92.38
[1.222, 1.528)	8	0.02	1.85	408	94.23
[1.528, 1.833)	6	0.01	1.39	414	95.61
[1.833, 2.139)	5	0.01	1.15	419	96.77
[2.139, 2.444)	8	0.02	1.85	427	98.61
[2.444, 2.75)	3	0.01	0.69	430	99.31
[2.75, 3.055)	3	0.01	0.69	433	100.00

Tabela 8: Óbitos em São Paulo

Classes c_i	frequência absoluta	frequência relativa	frequência relativa (%)	frequência acumulada	frequência acumulada (%)
[0, 9.6158)	96	0.22	22.17	96	22.17
[9.6158, 19.232)	76	0.18	17.55	172	39.72
[19.232, 28.847)	95	0.22	21.94	267	61.66
[28.847, 38.463)	74	0.17	17.09	341	78.75
[38.463, 48.079)	46	0.11	10.62	387	89.38
[48.079, 57.695)	22	0.05	5.08	409	94.46
[57.695, 67.311)	16	0.04	3.70	425	98.15
[67.311, 76.927)	5	0.01	1.15	430	99.31
[76.927, 86.542)	0	0.00	0.00	430	99.31
[86.542, 96.158)	3	0.01	0.69	433	100.00

Tabela 9: Casos em Espírito Santo

Classes c_i	frequência absoluta	frequência relativa	frequência relativa (%)	frequência acumulada	frequência acumulada (%)
[0, 0.279)	135	0.31	31.18	135	31.18
[0.279, 0.5579)	122	0.28	28.18	257	59.35
[0.5579, 0.8369)	98	0.23	22.63	355	81.99
[0.8369, 1.116)	34	0.08	7.85	389	89.84
[1.116, 1.395)	17	0.04	3.93	406	93.76
[1.395, 1.674)	9	0.02	2.08	415	95.84
[1.674, 1.953)	7	0.02	1.62	422	97.46
[1.953, 2.232)	4	0.01	0.92	426	98.38
[2.232, 2.511)	5	0.01	1.15	431	99.54
[2.511, 2.79)	2	0.00	0.46	433	100.00

Tabela 10: Óbitos em Espírito Santo

Classes c_i	frequência absoluta	frequência relativa	frequência relativa (%)	frequência acumulada	frequência acumulada (%)
[0, 4.9052)	164	0.38	37.88	164	37.88
[4.9052, 9.8105)	79	0.18	18.24	243	56.12
[9.8105, 14.716)	70	0.16	16.17	313	72.29
[14.716, 19.621)	49	0.11	11.32	362	83.60
[19.621, 24.526)	45	0.10	10.39	407	94.00
[24.526, 29.431)	16	0.04	3.70	423	97.69
[29.431, 34.337)	6	0.01	1.39	429	99.08
[34.337, 39.242)	3	0.01	0.69	432	99.77
[39.242, 44.147)	0	0.00	0.00	432	99.77
[44.147, 49.052)	1	0.00	0.23	433	100.00

Tabela 11: Casos em Rio de Janeiro

Classes c_i	frequência absoluta	frequência relativa	frequência relativa (%)	frequência acumulada	frequência acumulada (%)
[0, 0.2609)	146	0.34	33.72	146	33.72
[0.2609, 0.5218)	72	0.17	16.63	218	50.35
[0.5218, 0.7827)	76	0.18	17.55	294	67.90
[0.7827, 1.044)	57	0.13	13.16	351	81.06
[1.044, 1.305)	43	0.10	9.93	394	90.99
[1.305, 1.565)	17	0.04	3.93	411	94.92
[1.565, 1.826)	6	0.01	1.39	417	96.30
[1.826, 2.087)	8	0.02	1.85	425	98.15
[2.087, 2.348)	5	0.01	1.15	430	99.31
[2.348, 2.609)	3	0.01	0.69	433	100.00

Tabela 12: Óbitos em Rio de Janeiro

6 Código

Segue o código feito na linguagem de programação R.

```
#####
#                               Banco de Dados - COVID 19                               #
#                                                                                             #
#                               Analise                                                                                             #
#####

rm(list = ls(all = TRUE)) #limpa a memoria
setwd("~/Documentos/SME0211_INFERENCIA/Trabalho1")
#library(readr)
library(readxl)
library(fdth)
#library(data.table)

#-----
# Leitura de dados em arquivos
#(os arquivos deverao estar na pasta indicada)

dados = read_excel("Dados_Sudeste.xlsx")
View(dados)

#####
#-----
# Selecionando as variaveis de interesse da tabela
```

```

# Definindo escala: casos por milhao e mortes por milhao
dados$casosNovos_100mil =
  (dados$casosNovos*100000)/dados$populacaoTCU2019
dados$obitosNovos_100mil =
  (dados$obitosNovos*100000)/dados$populacaoTCU2019

# casos por dia
attach(dados)

casosNovosMG = subset(casosNovos_100mil, estado == 'MG')
casosNovosSP = subset(casosNovos_100mil, estado == 'SP')
casosNovosES = subset(casosNovos_100mil, estado == 'ES')
casosNovosRJ = subset(casosNovos_100mil, estado == 'RJ')

# obitos por dia
obitosNovosMG = subset(obitosNovos_100mil, estado == 'MG')
obitosNovosSP = subset(obitosNovos_100mil, estado == 'SP')
obitosNovosES = subset(obitosNovos_100mil, estado == 'ES')
obitosNovosRJ = subset(obitosNovos_100mil, estado == 'RJ')

n_MG = length(casosNovosMG)
n_SP = length(casosNovosSP)
n_ES = length(casosNovosES)
n_RJ = length(casosNovosRJ)

detach(dados)

#####
# item (b)

SP <- subset(dados, dados$estado == "SP")
RJ <- subset(dados, dados$estado == "RJ")
MG <- subset(dados, dados$estado == "MG")
ES <- subset(dados, dados$estado == "ES")
SP$casosAcumulado = SP$casosAcumulado*100000/SP$populacaoTCU2019
RJ$casosAcumulado = RJ$casosAcumulado*100000/RJ$populacaoTCU2019
MG$casosAcumulado = MG$casosAcumulado*100000/MG$populacaoTCU2019
ES$casosAcumulado = ES$casosAcumulado*100000/ES$populacaoTCU2019

SP$obitosAcumulado = SP$obitosAcumulado*100000/SP$populacaoTCU2019
RJ$obitosAcumulado = RJ$obitosAcumulado*100000/RJ$populacaoTCU2019
MG$obitosAcumulado = MG$obitosAcumulado*100000/MG$populacaoTCU2019
ES$obitosAcumulado = ES$obitosAcumulado*100000/ES$populacaoTCU2019

```

```

# Grafico Acumulado
x11()
plot(SP$casosAcumulado,
      xaxt = "n",
      xlab = "data", ylab = "Casos",
      col = ("red"),
      type = "l",
      main = "Numero_de_casos_diarios_Acumulado")

axis(side = 1,
      at = seq(1, n_SP, n_SP%%4),
      labels = SP$data[seq(1, n_SP, n_SP%%4)])

op <- par(new = TRUE)
plot(MG$casosAcumulado, type = "l", xlab = "", ylab = "",
      xaxt = "n", yaxt = "n", col = ("blue"))

op <- par(new = TRUE)
plot(RJ$casosAcumulado, type = "l", xlab = "", ylab = "",
      xaxt = "n", yaxt = "n", col = ("green"))

op <- par(new = TRUE)
plot(ES$casosAcumulado, type = "l", xlab = "", ylab = "",
      xaxt = "n", yaxt = "n", col = ("purple"))

legend("bottomright",
      col = c("red", "blue", "green", "purple"),
      legend = c("SP", "MG", "RJ", "ES"),
      lty = 1,
      lwd = 3,
      bty = "n")

op <- par(lty = 1, lwd = 1)

plot(SP$obitosAcumulado,
      xaxt = "n",
      xlab = "data", ylab = "obitos",
      col = ("red"),
      type = "l",
      main = "Numero_de_Obitos_diarios_Acumulado")

axis(side = 1,

```

```

    at = seq(1, n_SP, n_SP%%4),
    labels = SP$data[seq(1, n_SP, n_SP%%4)])

op <- par(new = TRUE)
plot(MG$obitosAcumulado, type = "l", xlab = "", ylab = "",
     xaxt = "n", yaxt = "n", col = ("blue"))

op <- par(new = TRUE)
plot(RJ$obitosAcumulado, type = "l", xlab = "", ylab = "",
     xaxt = "n", yaxt = "n", col = ("green"))

op <- par(new = TRUE)
plot(ES$obitosAcumulado, type = "l", xlab = "", ylab = "",
     xaxt = "n", yaxt = "n", col = ("purple"))

legend("bottomright",
      col = c("red", "blue", "green", "purple"),
      legend = c("SP", "MG", "RJ", "ES"),
      lty = 1,
      lwd = 3,
      bty = "n")

op <- par(lty = 1, lwd = 1)

# Graficos casos diarios // obitos diarios

x11()
plot(1:n_MG, casosNovosMG, type='b', pch=16, col='blue',
     cex=1.5, xlab='Dia', ylab='Casos_Novos_por_100.000_hab._(MG)')

x11()
plot(1:n_SP, casosNovosSP, type='b', pch=16, col='red',
     cex=1.5, xlab='Dia', ylab='Casos_Novos_por_100.000_hab._(SP)')

x11()
plot(1:n_ES, casosNovosES, type='b', pch=16, col='purple',
     cex=1.5, xlab='Dia', ylab='Casos_Novos_por_100.000_hab._(ES)')

x11()
plot(1:n_RJ, casosNovosRJ, type='b', pch=16, col='green',
     cex=1.5, xlab='Dia', ylab='Casos_Novos_por_100.000_hab._(RJ)')

```

```

x11()
plot(1:n_MG,obitosNovosMG, type='b', pch=16,col='blue',
      cex=1.5, xlab='Dia', ylab='Obitos_Novos_por_100.000_hab._(MG)')

x11()
plot(1:n_SP,obitosNovosSP, type='b', pch=16,col='red',
      cex=1.5, xlab='Dia', ylab='Obitos_Novos_por_100.000_hab._(SP)')

x11()
plot(1:n_ES,obitosNovosES, type='b', pch=16,col='purple',
      cex=1.5, xlab='Dia', ylab='Obitos_Novos_por_100.000_hab._(ES)')

x11()
plot(1:n_RJ,obitosNovosRJ, type='b', pch=16,col='green',
      cex=1.5, xlab='Dia', ylab='Obitos_Novos_por_100.000_hab._(RJ)')

#####
# item (c)

summary(casosNovosMG)
summary(obitosNovosMG)
summary(casosNovosSP)
summary(obitosNovosSP)
summary(casosNovosRJ)
summary(obitosNovosRJ)
summary(casosNovosES)
summary(obitosNovosES)

#####
# itens (d) e (e)

# MINAS GERAIS
# casos Novos
minCasosMG = min(casosNovosMG)
maxCasosMG = max(casosNovosMG)
AmplitudeCasosMG = maxCasosMG - minCasosMG
mediaCasosMG = mean(casosNovosMG)
medianaCasosMG = median(casosNovosMG)
quantilCasosMG = quantile(casosNovosMG, c(0.25,0.75), tipe=2)

```

```

AmplitudeQCasosMG = quantilCasosMG[1] - quantilCasosMG[2]
desvioPadraoCasosMG = sd(casosNovosMG)

# obitos Novos
minMortesMG = min(obitosNovosMG)
maxMortesMG = max(obitosNovosMG)
AmplitudeMortesMG = maxMortesMG - minMortesMG
mediaMortesMG = mean(obitosNovosMG)
medianaMortesMG = median(obitosNovosMG)
quantilMortesMG = quantile(obitosNovosMG, c(0.25,0.75), tipe=2)
AmplitudeQMortesMG = quantilMortesMG[1] - quantilMortesMG[2]
desvioPadraoMortesMG = sd(obitosNovosMG)

# Sao Paulo
# casos Novos
minCasosSP = min(casosNovosSP)
maxCasosSP = max(casosNovosSP)
AmplitudeCasosSP = - minCasosSP + maxCasosSP
mediaCasosSP = mean(casosNovosSP)
medianaCasosSP = median(casosNovosSP)
quantilCasosSP = quantile(casosNovosSP, c(0.25,0.75), tipe=2)
AmplitudeQCasosSP = quantilCasosSP[1] - quantilCasosSP[2]
desvioPadraoCasosSP = sd(casosNovosSP)

# obitos Novos
minMortesSP = min(obitosNovosSP)
maxMortesSP = max(obitosNovosSP)
AmplitudeMortesP = - minMortesSP + maxMortesSP
mediaMortesSP = mean(obitosNovosSP)
medianaMortesSP = median(obitosNovosSP)
quantilMortesSP = quantile(obitosNovosSP, c(0.25,0.75), tipe=2)
AmplitudeQMortesSP = quantilMortesSP[1] - quantilMortesSP[2]
desvioPadraoMortesSP = sd(obitosNovosSP)

# Rio de Janeiro
# casos Novos
minCasosRJ = min(casosNovosRJ)
maxCasosRJ = max(casosNovosRJ)
AmplitudeCasosRJ = - minCasosRJ + maxCasosRJ
mediaCasosRJ = mean(casosNovosRJ)
medianaCasosRJ = median(casosNovosRJ)
quantilCasosRJ = quantile(casosNovosRJ, c(0.25,0.75), tipe=2)
AmplitudeQCasosRJ = quantilCasosRJ[1] - quantilCasosRJ[2]

```



```

desvioPadraoCasosRJ = sd(casosNovosRJ)

# obitos Novos
minMortesRJ = min(obitosNovosRJ)
maxMortesRJ = max(obitosNovosRJ)
AmplitudeMortesRJ = - minMortesRJ + maxMortesRJ
mediaMortesRJ = mean(obitosNovosRJ)
medianaMortesRJ = median(obitosNovosRJ)
quantilMortesRJ = quantile(obitosNovosRJ, c(0.25,0.75), tipe=2)
AmplitudeQMortesRJ = quantilMortesRJ[1] - quantilMortesRJ[2]
desvioPadraoMortesRJ = sd(obitosNovosRJ)

# Espirito Santo
# casos Novos
minCasosES = min(casosNovosES)
maxCasosES = max(casosNovosES)
AmplitudeCasosES = - minCasosES + maxCasosES
mediaCasosES = mean(casosNovosES)
medianaCasosES = median(casosNovosES)
quantilCasosES = quantile(casosNovosES, c(0.25,0.75), tipe=2)
AmplitudeQCasosES = quantilCasosES[1] - quantilCasosES[2]
desvioPadraoCasosES = sd(casosNovosES)

# obitos Novos
minMortesES = min(obitosNovosES)
maxMortesES = max(obitosNovosES)
AmplitudeMortesES = - minMortesES + maxMortesES
mediaMortesES = mean(obitosNovosES)
medianaMortesES = median(obitosNovosES)
quantilMortesES = quantile(obitosNovosES, c(0.25,0.75), tipe=2)
AmplitudeQMortesES = quantilMortesES[1] - quantilMortesES[2]
desvioPadraoMortesES = sd(obitosNovosES)

#####
# item (f)

# Histograma (amplitude das classes eh dada pela regra de Sturges)

# Minas Gerais
hist(casosNovosMG,
      xlab = "Numero de casos",

```

```

        ylab = "Frequencia",
        col = "blue",
        main = "Histograma_de_casos_diarios_registrados_em_MG")
hist(obitosNovosMG,
     xlab = "Numero_de_obitos",
     ylab = "Frequencia",
     col = "blue",
     main = "Histograma_de_obitos_diarios_registrados_em_MG")

# Sao Paulo
hist(casosNovosSP,
     xlab = "Numero_de_casos",
     ylab = "Frequencia",
     col = "red",
     main = "Histograma_de_casos_diarios_registrados_em_SP")
hist(obitosNovosSP,
     xlab = "Numero_de_obitos",
     ylab = "Frequencia",
     col = "red",
     main = "Histograma_de_obitos_diarios_registrados_em_SP")

# Rio de Janeiro
hist(casosNovosRJ,
     xlab = "Numero_de_casos",
     ylab = "Frequencia",
     col = "green",
     main = "Histograma_de_casos_diarios_registrados_em_RJ")
hist(obitosNovosRJ,
     xlab = "Numero_de_obitos",
     ylab = "Frequencia",
     col = "green",
     main = "Histograma_de_obitos_diarios_registrados_em_RJ")

# Espirito Santo
hist(casosNovosES,
     xlab = "Numero_de_casos",
     ylab = "Frequencia",
     col = "purple",
     main = "Histograma_de_casos_diarios_registrados_em_ES")
hist(obitosNovosES,
     xlab = "Numero_de_obitos",
     ylab = "Frequencia",

```

```

        col = "purple",
        main = "Histograma de obitos diários registrados em ES")

# Boxplot
# Casos Novos
casos = data.frame(MG = casosNovosMG,
                   SP = casosNovosSP,
                   ES = casosNovosES,
                   RJ = casosNovosRJ)

boxplot(casos,
        ylab = "Numero de casos",
        main = "Boxplot do numero de casos diários",
        col = c("blue", "red", "purple", "green"))

# Obitos Novos
obitos = data.frame(MG = obitosNovosMG,
                   SP = obitosNovosSP,
                   ES = obitosNovosES,
                   RJ = obitosNovosRJ)

boxplot(obitos,
        ylab = "Numero de Obitos",
        main = "Boxplot do numero de obitos diários",
        col = c("blue", "red", "purple", "green"))

#####

```