

lista1_gabarito

Gabriel_Mardegan

2024-08-22

Métodos quantitativos - regressão linear

Lista 1 - revisão estatística e introdução ao uso do R [gabarito]

monitor: Gabriel Rodrigues Mardegan

1. Importando a base de dados

1.1) A unidade de análise do banco de dados são os municípios brasileiros no ano de 2020. O número de observações (municípios) listadas é: 5570.

1.2) A base de dados importada possui 8 variáveis.

1.3) [pergunta em aberto] As colunas/variáveis “populacao_str” e “populacao” apresentam os mesmos números para a população dos municípios. No entanto, na primeira os valores estão no tipo “character”, e na segunda no tipo “numeric”. Isso é possível verificar posicionando o cursor em cima do nome da coluna, o que também fornece informação sobre a variação dos valores para essa variável. Em relação ao conteúdo da base de dados, é curioso notar como os municípios tem uma variação considerável em relação à população. Uma tabela como essa permite entender quantos municípios brasileiros (porcentagem) são de pequeno, médio ou grande porte, o que direciona formulação e execução de políticas públicas.

2. Analisando o estado de São Paulo

2.1) De acordo com a base de dados do IBGE para o ano de 2020, o estado de São Paulo possui 645 municípios.

2.2) O menor município do estado no ano de 2020 foi borá, com uma população de 838 habitantes.

2.3) Os municípios paulistas com população superior a um milhão de habitantes são: Campinas (1.213.792 habitantes), Guarulhos (1.392.121 habitantes) e São Paulo (12.325.232 habitantes).

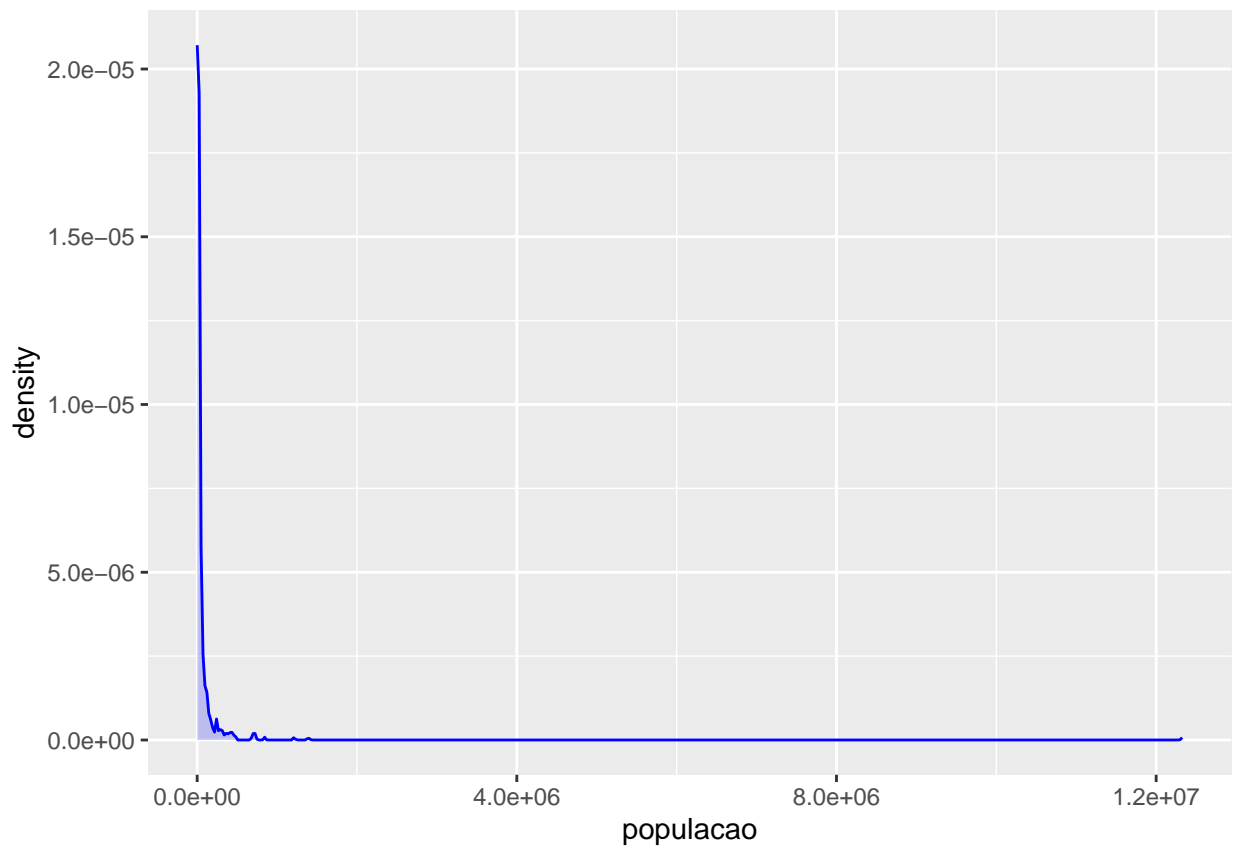
3. Estatística descritiva dos dados

Table 1: estatísticas descritivas da população paulista 2020

media	mediana	desvio	variancia
71766.41	14141	498489.9	248492160624

4. Trabalhando com gráficos

4.1)

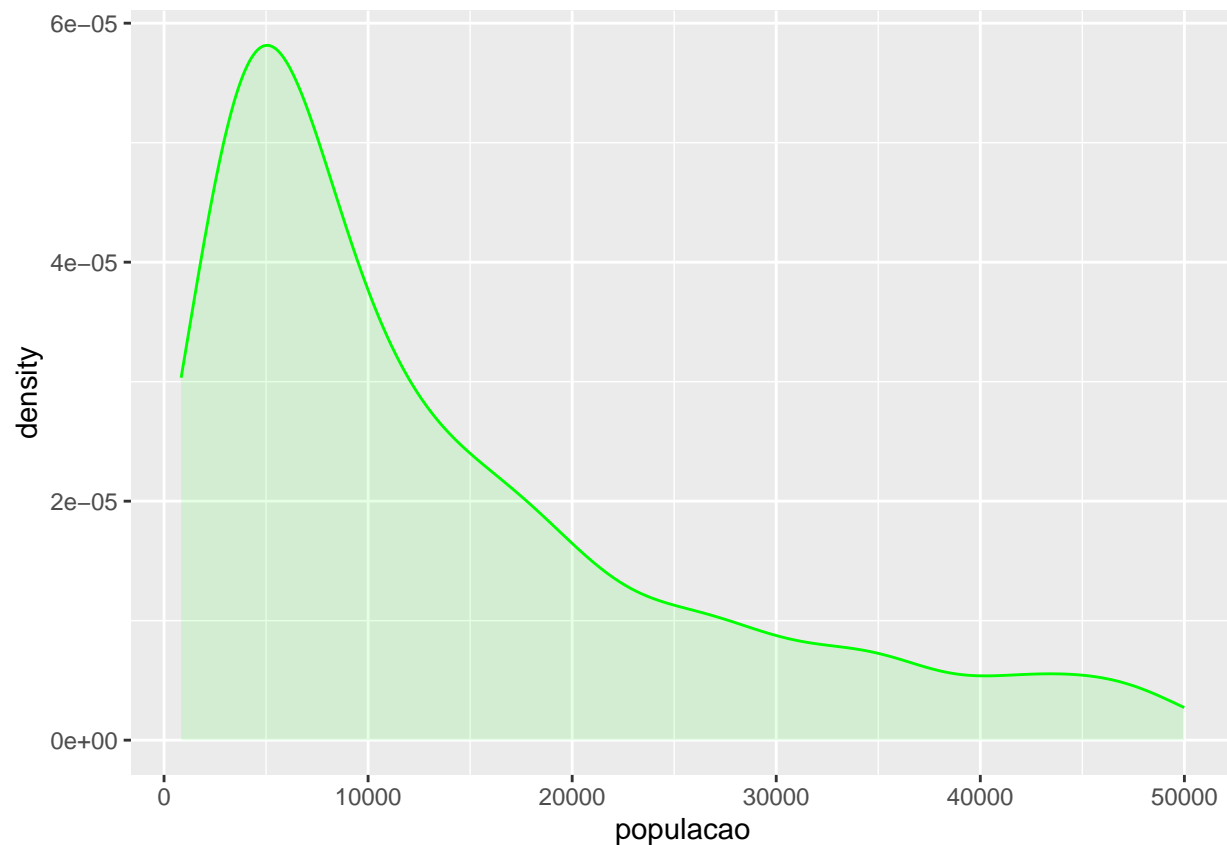


4.2) O gráfico de densidade criado revela que a maioria dos municípios paulistas estão concentrados na região próxima a zero. A existência de um município com uma população muito maior que o restante (valor discrepante extremo), a capital de São Paulo, faz com que a cauda direita se alongue até o ponto de 12 milhões de habitantes, dificultando a visualização da distribuição dos municípios.

4.3) A partir da observação do gráfico, constata-se que a mediana é uma medida mais adequada da tendência central da população, uma vez que a distribuição é muito assimétrica.

5. Gráfico de densidade de pequenos municípios

5.1)



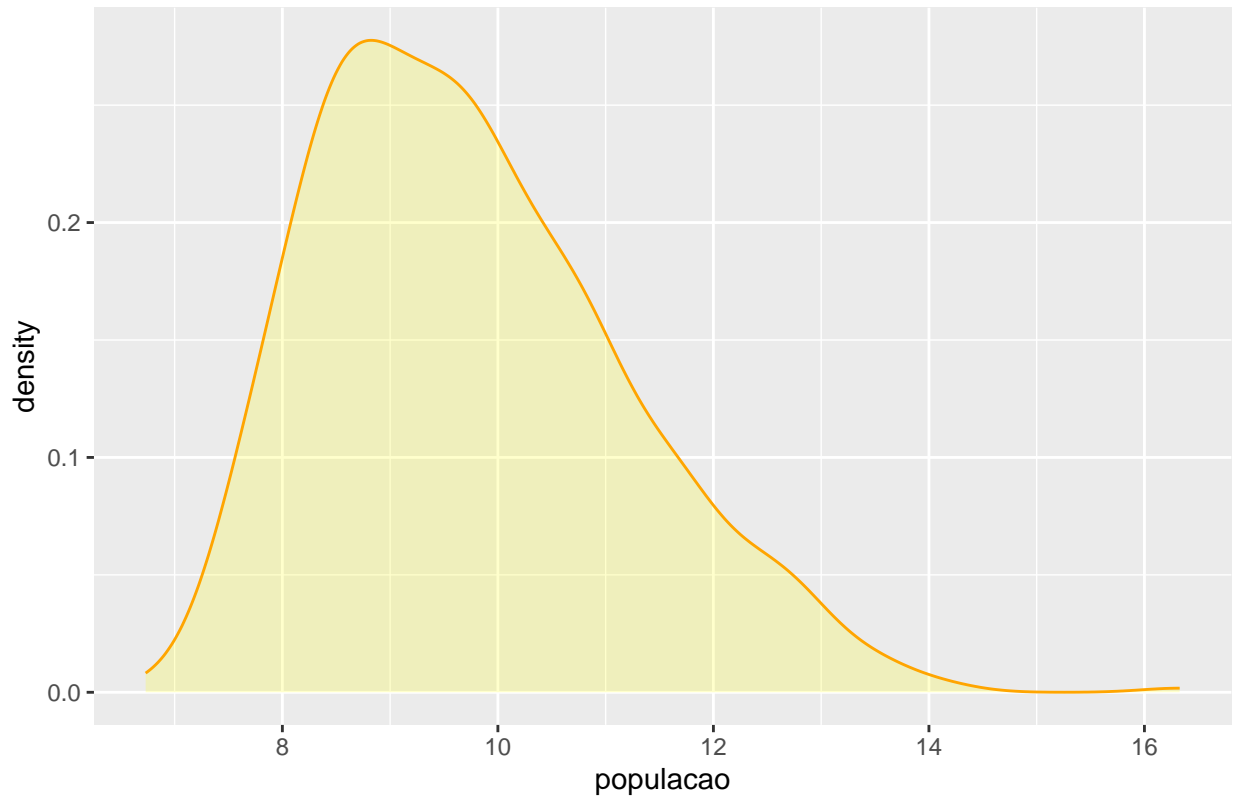
5.2) O estado de São Paulo possui 504 municípios com menos de 50.000 habitantes (para o ano de 2020). Isso representa 78.1395349% dos municípios paulistas.

5.3) Nesse novo gráfico, as observações/municípios não estão mais concentrados em apenas uma parte do gráfico. O gráfico ainda é assimétrico, mas bem menos que o anterior.

6. Trabalhando com escala logarítmica

6.1)

Grafico de densidade da populacao (em log)



6.2) Em comparação com o gráfico do exercício 4.1, este gráfico tem uma aparência muito mais simétrica, e as observações não estão concentradas em apenas um lado do gráfico. Apesar disso, a cauda direita ainda se estende, produzindo uma assimetria, mas muito diferente do gráfico anterior. Como visto em aula, transformar dados em escala logarítmica permite melhor visualização de dados com ordem de grandeza diferentes (10, 100, 1000...).

Médias da população por estado

7.1)

Table 2: medias populacionais dos municipios por estado

uf	pop_mean
AC	40657.73
AL	32858.26
AM	67866.35
AP	53860.81
BA	35804.88
CE	49929.91
DF	3055149.00
ES	52103.23

uf	pop_mean
GO	28916.83
MA	32786.17
MG	24962.09
MS	35561.95
MT	25008.65
PA	60352.40
PB	18113.35
PE	51981.74
PI	14649.46
PR	28864.26
RJ	188762.92
RN	21162.66
RO	34547.31
RR	42078.73
RS	22983.85
SC	24584.75
SE	30917.63
SP	71766.41
TO	11440.63

7.2)

Table 3: tabela ordenada de médias dos municípios por estado

uf	pop_mean
TO	11440.63
PI	14649.46
PB	18113.35
RN	21162.66
RS	22983.85
SC	24584.75
MG	24962.09
MT	25008.65
PR	28864.26
GO	28916.83
SE	30917.63
MA	32786.17
AL	32858.26
RO	34547.31
MS	35561.95
BA	35804.88

uf	pop_mean
AC	40657.73
RR	42078.73
CE	49929.91
PE	51981.74
ES	52103.23
AP	53860.81
PA	60352.40
AM	67866.35
SP	71766.41
RJ	188762.92
DF	3055149.00

Tocantins possui a menor média, e Rio de Janeiro tem a maior média (desconsiderando DF).