

MAE5763 - Modelos Lineares Generalizados - Resolução da Lista 1

Guilherme Marthe - 8661962

9/21/2020

Exercício 1

em progresso

Exercício 2

em progresso

Exercício 3

em progresso

Exercício 4

em progresso

Exercício 5

em progresso

Exercício 6

A seguir iremos analisar a base **fuel2001.txt**. Conforme o enunciado, são descritas as seguintes variáveis referentes aos 50 estados norte-americanos mais o Distrito de Columbia no ano de 2001:

- i) UF, unidade da federação
- ii) Drivers, número de motoristas licenciados
- iii) FuelC, total de gasolina vendida (em mil galões)
- iv) Income, renda per capita em 2000 (em mil USD)
- v) Miles, total de milhas em estradas federais
- vi) MPC, milhas per capita percorridas
- vii) Pop, população ≥ 16 anos
- viii) Tax, taxa da gasolina (em cents por galão)

Abaixo mostro as 6 primeiras linhas da base.

| State | Drivers | FuelC | Income | Miles | MPC | Pop | Tax |
|-------|----------|----------|--------|--------|----------|----------|------|
| AL | 3559897 | 2382507 | 23471 | 94440 | 12737.00 | 3451586 | 18.0 |
| AK | 472211 | 235400 | 30064 | 13628 | 7639.16 | 457728 | 8.0 |
| AZ | 3550367 | 2428430 | 25578 | 55245 | 9411.55 | 3907526 | 18.0 |
| AR | 1961883 | 1358174 | 22257 | 98132 | 11268.40 | 2072622 | 21.7 |
| CA | 21623793 | 14691753 | 32275 | 168771 | 8923.89 | 25599275 | 18.0 |
| CO | 3287922 | 2048664 | 32949 | 85854 | 9722.73 | 3322455 | 22.0 |

Conforme sugerido, realizaremos as transformações $Fuel = 1000FuelC/Pop$ e $Dlic = 1000Drivers/Pop$ para possibilitar a comparação entre estados com diferentes populações. Em seguida criamos a variável $lMiles = \log(Miles)$.

Com essas transformações, partimos para uma análise descritiva da base, antes de partirmos para a modelagem.

Análise univariada

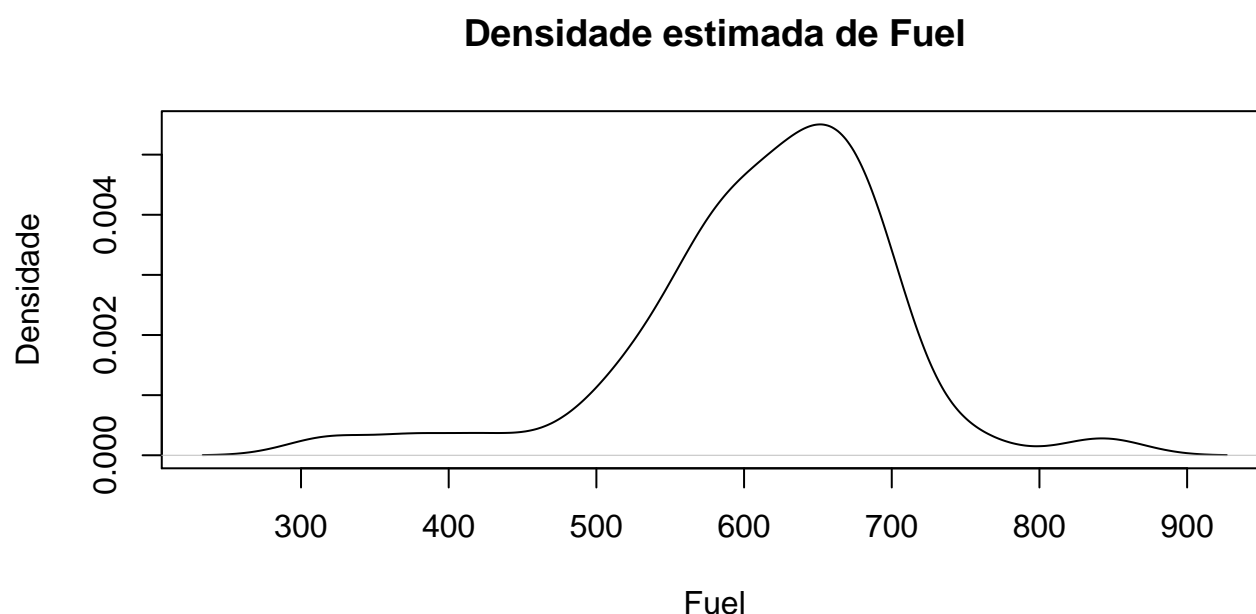
Primeiramente, a partir de algumas estatísticas descritivas de cada uma das variáveis estudadas que mostro abaixo, podemos postular algumas características das variáveis:

- a variável *fuel* mostra uma leve assimetria à esquerda uma vez que sua média é menor que a mediana
- a variável *income* mostra uma assimetria à direita uma vez que sua média é maior que a mediana

| variável | média | desvio padrão | quartil 1 | mediana | quartil 3 |
|----------|-------|---------------|-----------|---------|-----------|
| dlic | 904 | 73 | 864 | 909 | 943 |
| fuel | 613 | 89 | 575 | 626 | 667 |
| income | 28404 | 4452 | 25323 | 27871 | 31208 |
| lmiles | 11 | 1 | 11 | 11 | 12 |
| tax | 20 | 5 | 18 | 20 | 23 |

Neste exercício, a variável *fuel* é a variável resposta. Por isso à inspecionaremos primeiro. o gráfico abaixo mostra a densidade estimada dela. Podemos ver que ela possui uma tendência gaussiana, apesar de possuir caudas ligeiramente pesadas.

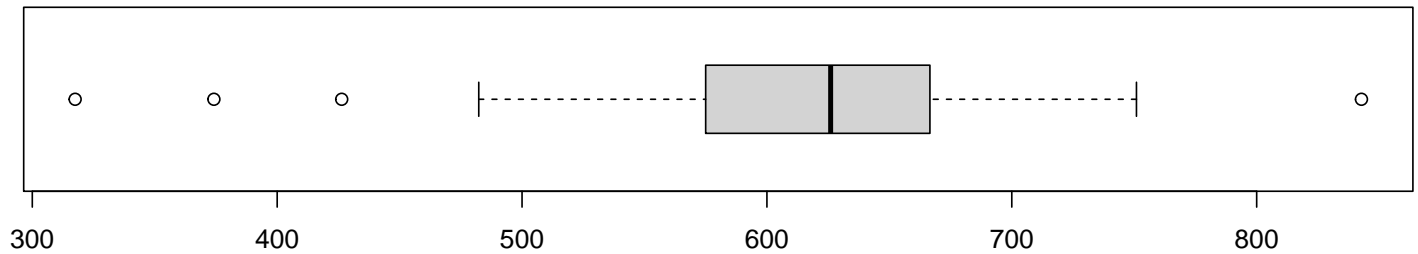
```
plot(density(mdata$fuel), xlab="Fuel", ylab="Densidade", main="Densidade estimada de Fuel")
```



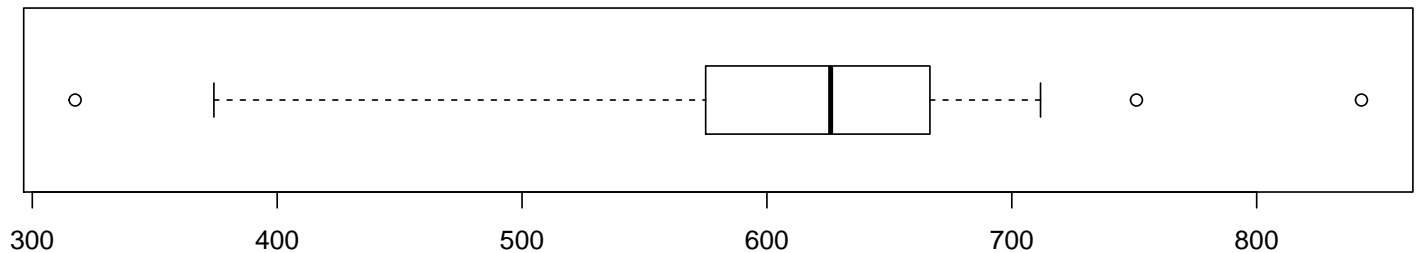
Abaixo mostramos os boxplots comum (criado através das estatísticas usuais como mediana e quartis) e robusto para a variável *fuel*. No boxplot comum estão 3 pontos que poderiam ser considerados extremos à esquerda, sendo mais baixo que o esperado para essa variável, e a extremo à direita. Com o boxplot robusto os pontos extremos à esquerda não são tão evidenciados, porém à direita um ponto novo aparece.

```
par(mfrow=c(2, 1))
boxplot(mdata$fuel, horizontal = T, main='Fuel - boxplot comum')
adjbox(mdata$fuel, horizontal = T, main='Fuel - boxplot robusto')
```

Fuel – boxplot comum

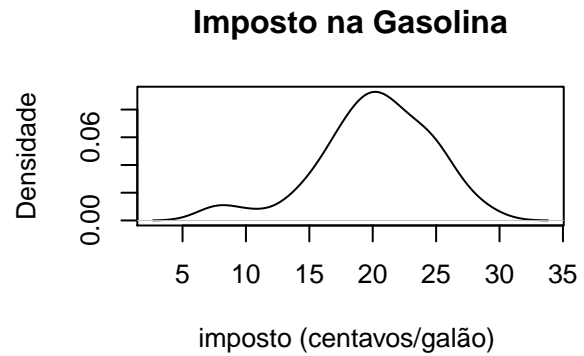
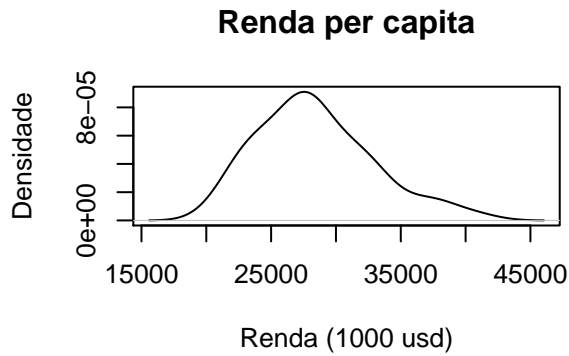
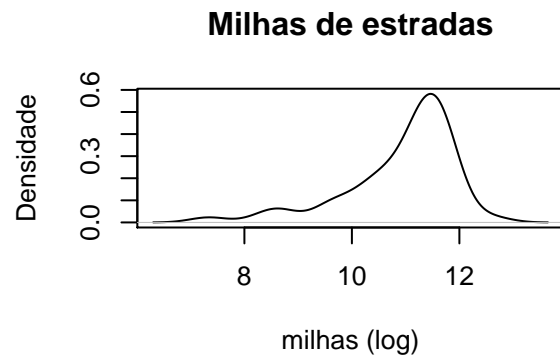
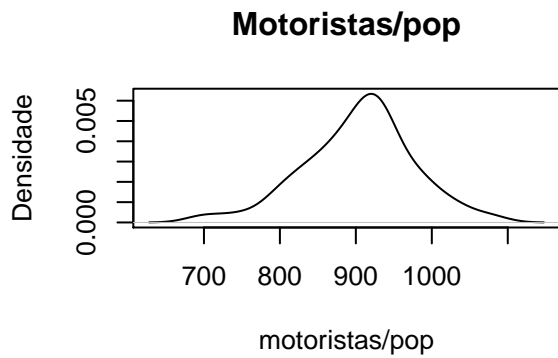


Fuel – boxplot robusto



Por completude, mostro as densidades empíricas das variáveis explicativas do estudo. Como sob a construção de modelos elas são consideradas fixas, suas distribuições não são importantes para o ajuste de modelos de regressão linear múltipla.

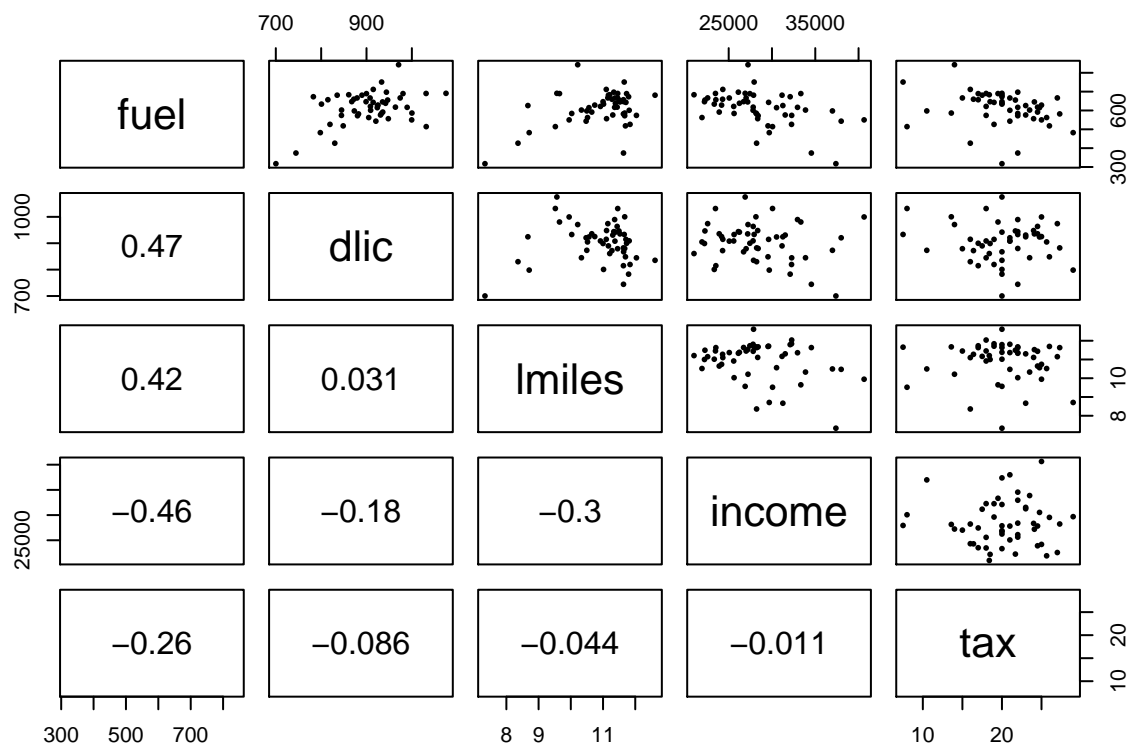
```
par(mfrow=c(2,2))
plot(density(mdata$dlic), xlab="motoristas/pop", ylab="Densidade", main='Motoristas/pop')
plot(density(mdata$lmliles), xlab="milhas (log)", ylab="Densidade", main='Milhas de estradas')
plot(density(mdata$income), xlab="Renda (1000 usd)", ylab="Densidade", main='Renda per capita')
plot(density(mdata$tax), xlab="imposto (centavos/galão) ", ylab="Densidade",
      main='Imposto na Gasolina')
```



Análise bivariada

```
panel.cor <- function(x, y, ...)
{
  par(usr = c(0, 1, 0, 1))
  txt <- as.character(format(cor(x, y), digits=2))
  text(0.5, 0.5, txt, cex = 1.57)
}

pairs(mdata, pch=16, cex=0.6, lower.panel = panel.cor)
```



Exercício 7

em progresso