

Bônus - Prova 3

Mariana Costa Freitas

2025-03-13

Introdução

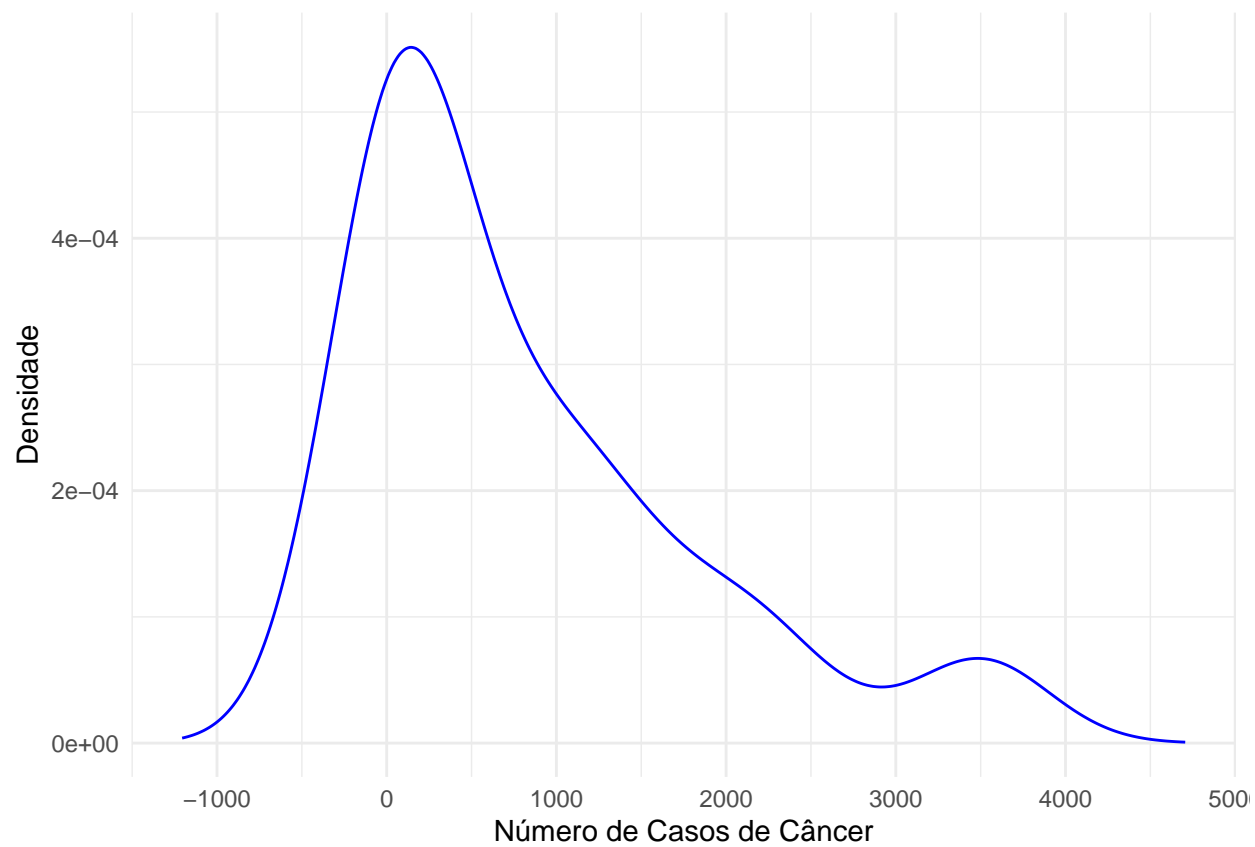
Metodologia

Análise dos dados

Análise Descritiva

Table 1: Resumo estatístico do dataset ccancer

count	gender	region	site	population
Min. : 0.00	F:15	Newfoundland:10	Breast :6	Min. : 533800
1st Qu.: 31.25	M:15	Ontario :10	Colorectal:6	1st Qu.: 533800
Median : 400.00	NA	Quebec :10	Lung :6	Median : 7410500
Mean : 814.83	NA	NA	Pancreas :6	Mean : 6606233
3rd Qu.:1212.50	NA	NA	Prostate :6	3rd Qu.:11874400
Max. :3500.00	NA	NA	NA	Max. :11874400



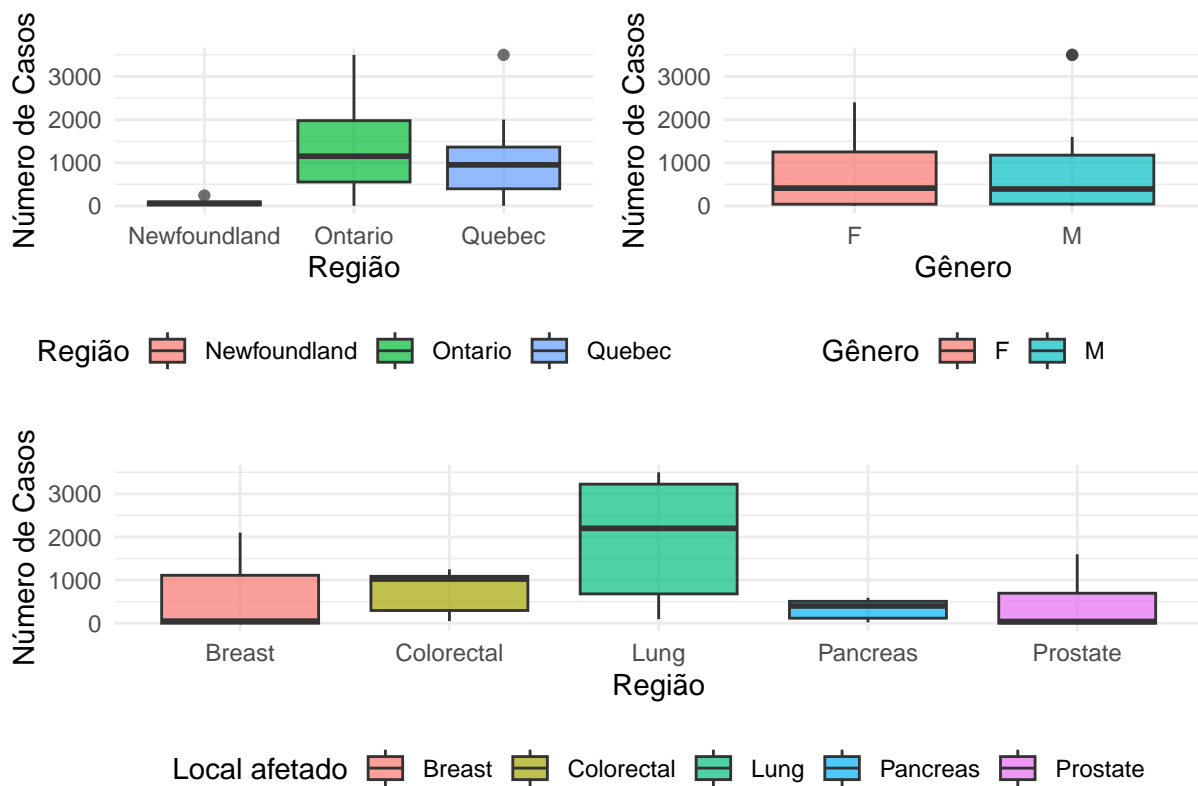


Table 2: Média e Variância da variável 'mpg'

Estatística	Valor
Média	814.8333
Variância	1,039,585.3161

Ajuste do modelo

A variável `population` não foi incluída no modelo, por ter variância extramente baixa

```
## [1] 9375.131
```

```
## [1] 21.95111
```

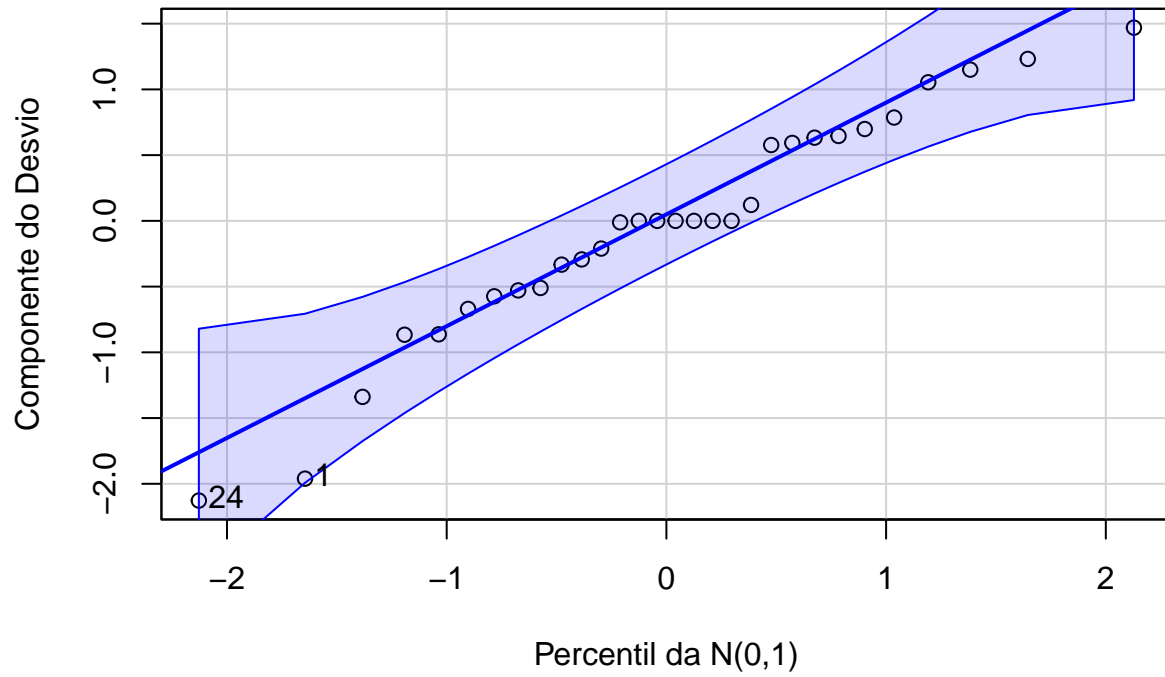
A função de ligação log é a mais comum e apropriada para modelos de contagem, como o binomial negativo, pois garante que os valores preditos sejam sempre positivos, o que faz sentido para variáveis de contagem. A função de ligação logit não é apropriada para variáveis de contagem, pois assume que a variável resposta está no intervalo (0, 1). A função de ligação probit é usada para respostas binárias ou proporcionais, assim como o logit. A função de ligação Complementay Log Log apresenta os mesmos problemas que a logit e probit. As funções de ligação identidade e raiz quadrática podem ser usadas para dados de contagem, porém não encontraram uma combinação de coeficientes válidos para ajuste do modelo. Assim, a única alternativa foi a logarítmica.

Population foi removida nos modelos, pois apresenta colinearidade, ou seja, essa é altamente correlacionada com outras variáveis explicativas ou mesmo com a variável resposta.

```
## Likelihood ratio tests of Negative Binomial Models
##
## Response: count
##
##           Model          theta Resid. df    2 x log-lik.
## 1              1          0.2700664      29      -418.0826
## 2             gender          0.2704422      28      -418.0327
## 3   gender + region + site 8740569.6998232      22      -9565.9311
## 4 gender + region + site + population 8740569.6998232      22      -9565.9311
## 5 gender * site + region + population 117.2354409      18      -255.7487
##
##      Test      df      LR stat.    Pr(Chi)
## 1
## 2 1 vs 2      1      0.04989336 0.8232489
## 3 2 vs 3      6 -9147.89831611 1.0000000
## 4 3 vs 4      0      0.00000000 1.0000000
## 5 4 vs 5      4  9310.18233215 0.0000000

##           (Intercept)          genderM          siteColorectal
##           4.5554301          -41.7105284          -0.5471315
##           siteLung          sitePancreas          siteProstate
##           0.1828127          -1.2723827          -43.5646545
##           regionOntario          regionQuebec          population
##           3.0335374          2.7887806          NA
## genderM:siteColorectal          genderM:siteLung          genderM:sitePancreas
##           41.8492504          42.2888149          41.6270906
## genderM:siteProstate
##           84.9480871
```

Análise de resíduos



[1] 24 1

