

RESUMO

Relatório de interpretação estatística dos dados relativos aos Estados Unidos da América com o objetivo de explicar como a expectativa de vida é influenciada por diversos fatores.

Palavras-chave: expectativa de vida, Estados Unidos, estatística.

Carlos Eduardo Fontaneli

carlos.fontaneli@estudante.ufscar.br

Hugo da Silva e Souza

hugo.souza@estudante.ufscar.br

Luana de Queiroz Garcia

luanagg@estudante.ufscar.br

Luís Augusto Simas do Nascimento

luissimas@estudante.ufscar.br

Matheus Rezende Milani Videira

matheus.videira@estudante.ufscar.br

Rafael Vinicius Polato Passador

rafaelpassador@estudante.ufscar.br

RELATÓRIO DE PROBABILIDADE E ESTATÍSTICA. *Análise Estatística*

1 INTRODUÇÃO

No conjunto de dados são apresentados os nomes de 50 estados norte-americanos, juntamente com as seguintes variáveis: *pop* (população estimada em julho de 1975), *percap* (renda per capita do ano em 1974 em USD), *analf* (proporção de analfabetos em 1970), *expvida* (expectativa de vida em anos de 1969-70), *crime* (taxa de criminalidade por 100.000 habitantes em 1976), *estud* (porcentagem de estudantes que concluem o segundo grau em 1970), *ndias* (número de dias do ano com temperatura abaixo de zero grau Celsius na cidade mais importante do estado) e *area* (área do estado em milhas quadradas). Ademais, o interesse neste estudo é tentar explicar a expectativa de vida da população a partir de análises das variáveis *percap*, *analf*, *crime*, *estud*, *ndias* e *dens* em que $dens = pop/area$.

Dessarte, a fim de realizar tais análises, será feito dois tipos de análises a partir do conjunto de dados fornecido, sendo elas: análise descritiva unidimensional e análise descritiva bidimensional. Haja vista que todas as variáveis disponibilizadas pela amostra são quantitativas, serão utilizadas, então, tabelas de frequências unidimensionais e tabelas de frequências bidimensionais, com suas respectivas tabelas de medidas descritivas correspondentes.

1.1 Amostra

Para realização desse experimento, o plano probabilístico e o processo de amostragem utilizados, os quais reúnem todas as técnicas que usam mecanismos aleatórios(sorteio) de seleção dos elementos da amostra, atribuindo a cada um deles uma probabilidade conhecida de seleção, foi a Amostragem Casual Simples (A.C.S), em que todos os elementos da população têm igual probabilidade de pertencer a amostra.

Considerou-se para o presente experimento as seguintes características, com seus respectivos tipos, dispostas na seguinte tabela:

Variáveis	Tipo de Variável	Descrição
<i>pop</i>	Quantitativa Contínua	Quantidade estimada de pessoas no estado
<i>percap</i>	Quantitativa Contínua	Renda per capita mensurada em dólares (USD)
<i>analf</i>	Quantitativa Contínua	Proporção de pessoas analfabetas em relação à população
<i>expvida</i>	Quantitativa Contínua	Expectativa de vida em anos
<i>crime</i>	Quantitativa Contínua	Taxa de criminalidade por 100.000 habitantes
<i>estud</i>	Quantitativa Contínua	Porcentagem de estudantes que concluem o ensino médio
<i>ndias</i>	Quantitativa Discreta	Quantidade de dias com a temperatura abaixo de 0 graus Celsius na cidade mais importante
<i>area</i>	Quantitativa Contínua	Área do estado em milhas quadradas
<i>dens</i>	Quantitativa Contínua	População dividida pela área do estado

1.2 Objetivos

Esta análise estatística possui como principal objetivo analisar e buscar uma explicação sobre a expectativa de vida nas presentes populações com bases nos dados relativos à renda per capita, ao índice de analfabetismo, à taxa de criminalidade, ao nível de estudo, ao número de dias com temperatura abaixo de zero e à densidade demográfica, expondo uma possível correlação entre as variáveis em questão.

1.3 Metodologia Estatística e Apresentação dos Resultados

A partir da organização, representação e sumarização dos dados obtidos pela análise da amostra, faz-se necessário obter maiores informações através de técnicas apropriadas. Nesse contexto, utilizou-se o software R para realizar a análise unidimensional e bidimensional dos dados, juntamente com alguns pacotes adicionais de funções da linguagem.

Nesse íterim, a fim de conhecer o comportamento de cada variável, analisou-se suas ocorrências através de tabelas de frequências, absolutas e relativas, bem como seus gráficos correspondentes, que ofertam uma ideia de como elas se distribuem perante a população. Assim sendo, as variáveis quantitativas discretas são representadas a partir da ocorrência do cópulus, concomitantemente, os dados relativos às variáveis quantitativas contínuas são agrupados em classes (faixas) escolhidas de maneira arbitrária pelo grupo.

Paralelamente, utilizou-se de gráficos para representar a distribuição das variáveis, bem como informações sobre variabilidade, simetria, valores discrepantes, entre outros aspectos

relevantes. Por conseguinte, como todas em questão são quantitativas, optou-se por representá-las através de histogramas.

Os dados categóricos são apresentados em números absolutos e percentuais, e as variáveis quantitativas em média e mediana e interquartis. A análise de associação entre as variáveis foi realizada a partir de análises gráficas e, concomitantemente, da correlação empírica entre estas. Ademais, para traçar as correlações empíricas entre as variáveis utilizou-se o coeficiente de correlação de Pearson.

Por fim, a fim de estimar correlações lineares entre as variáveis realizou-se uma análise de regressão linear. Dessarte, testes de hipóteses para os coeficientes de correlação, representações gráficas e verificação de adequabilidade das hipóteses estabelecidas sobre normalidade dos erros e variância constante foram realizadas também, em busca de sustentar a análise de regressão linear feita.

2 ANÁLISE DESCRITIVA UNIDIMENSIONAL

I. População:

Dos 50 estados avaliados, 22 (44,9%) estados têm uma população menor que 2.448 habitantes e somente 2 estados (2,04%) tem população maior que 12.864 habitantes. Ademais, a média de habitantes por estado é de 4.246 habitantes.

Tabela 1.1. Tabela de frequência relativa à variável quantitativa *população*.

População

Tabela de frequências

Quantidade	Absoluta	Relativa %
[365, 2448.3[22	44
[2448.3, 4531.6[12	24
[4531.6, 6614.9[7	14
[6614.9, 8698.2[2	4
[8698.2, 10781.5[2	4
[10781.5, 12864.8[3	6
[12864.8, 14948.1[0	0
[14948.1, 17031.4[0	0
[17031.4, 19114.7[1	2
[19114.7, 21198[1	2

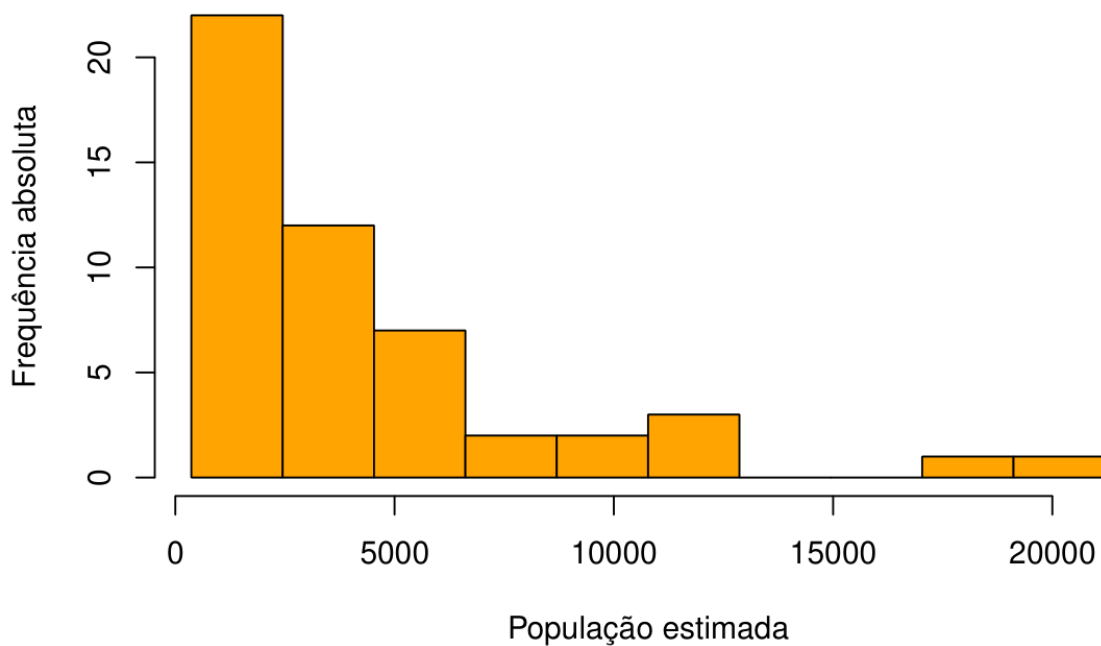


Gráfico 1.1. Histograma relativo à variável quantitativa contínua *população*.

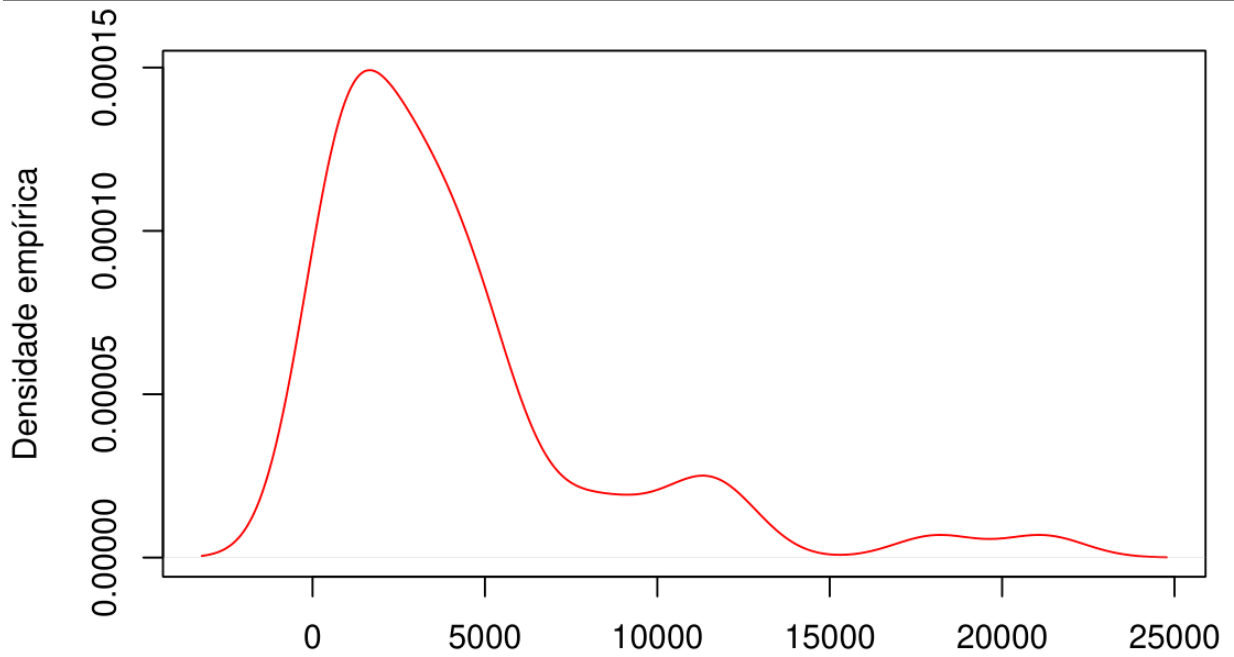


Gráfico 1.2. Gráfico de densidade empírica relativo à variável quantitativa contínua *população*.

Tabela 1.2. Medidas descritivas relativas à variável quantitativa contínua *população*.

Medida	Valor
Min.	365.00
1st Qu.	1079.50
Median	2838.50
Mean	4246.42
3rd Qu.	4968.50
Max.	21198.00

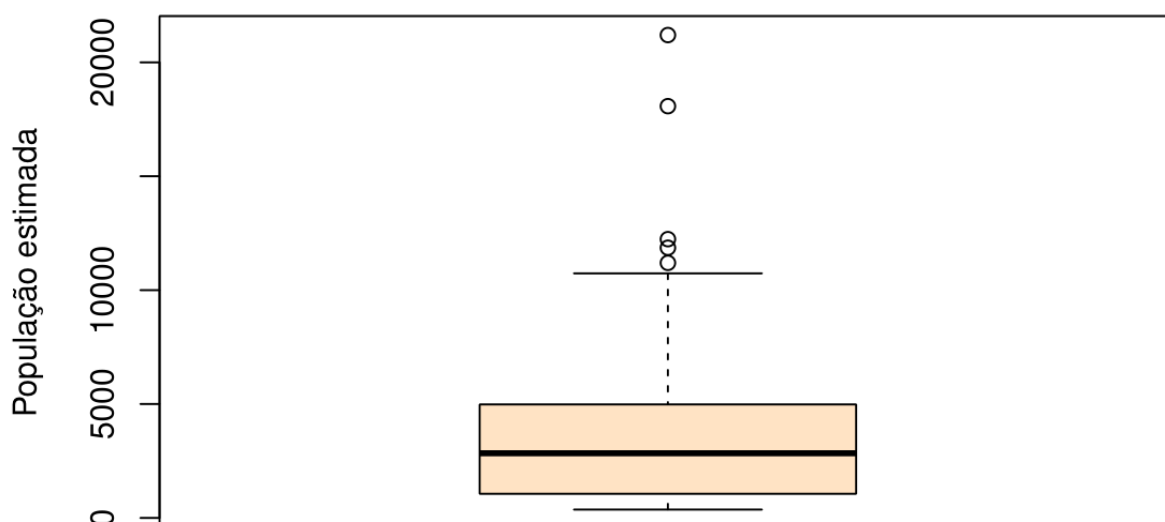


Figura 1. Boxplot relativo à variável quantitativa contínua *população*.

II. Renda per capita:

Dos 50 estados avaliados, 18(36%) deles possuem uma renda per capita entre R\$4500 e \$5000, 2(4%) estados possuem entre \$3000,00 e \$3500,00 e somente 1 (2%) possuem renda acima de \$6000,00. A renda per capita média entre os estados é de \$4435,80, sendo a mínima \$3098,00 (Mississippi) e a máxima \$6315,00 (Alaska).

Tabela 2.1. Tabela de frequência relativa à variável quantitativa *renda per capita*.

Renda per capita

Tabela de frequências

Valores	Absoluta	Relativa %
[\$3000, \$3500[2	4
[\$3500, \$4000[11	22
[\$4000, \$4500[11	22
[\$4500, \$5000[18	36
[\$5000, \$5500[7	14
[\$5500, \$6000[0	0
[\$6000, \$6500[1	2

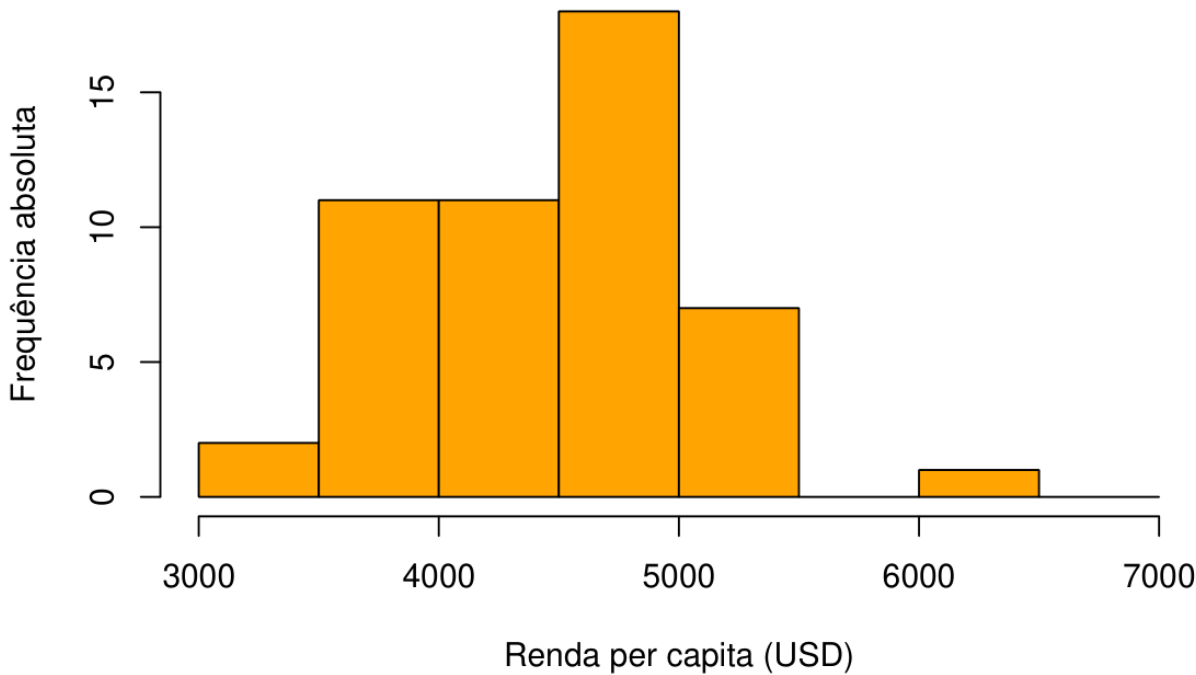


Gráfico 2.1. Histograma relativo à variável quantitativa contínua *renda per capita*.

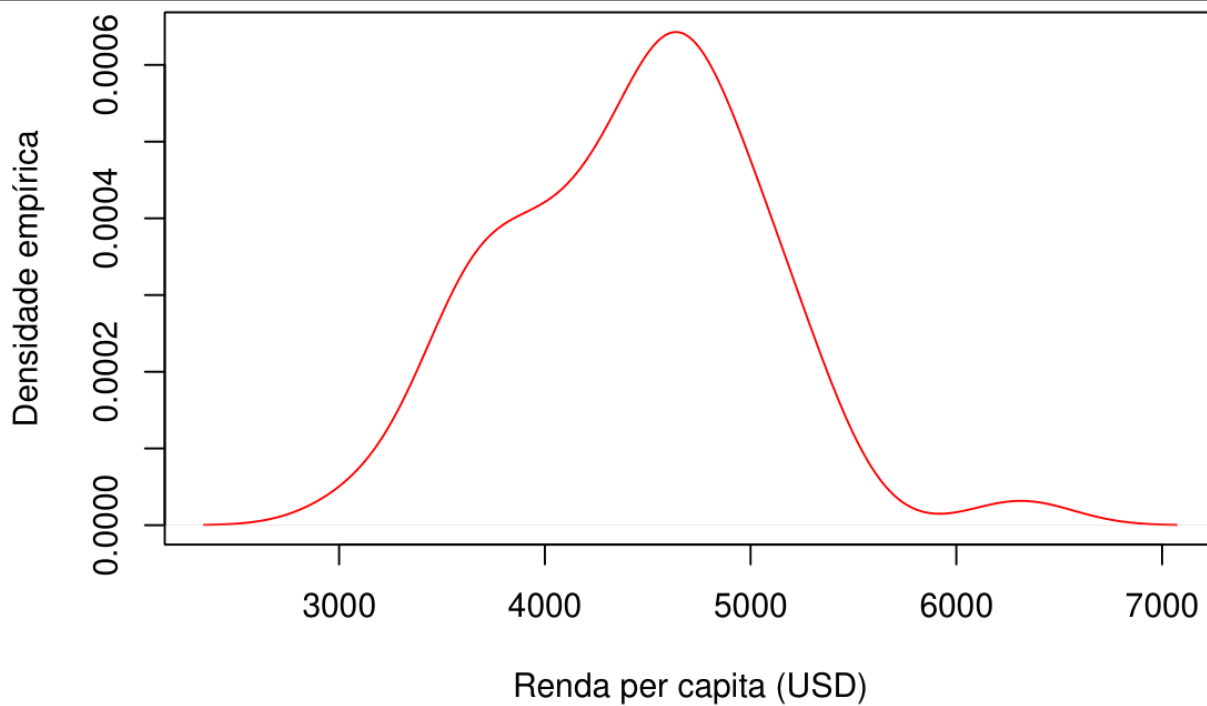


Gráfico 2.2. Gráfico de densidade empírica relativo à variável quantitativa contínua *renda per capita*.

Tabela 2.2. Medidas descritivas relativas à variável quantitativa contínua *renda per capita*.

Medida	Valor
Min.	3098.00
1st Qu.	3992.75
Median	4519.00
Mean	4435.80
3rd Qu.	4813.50
Max.	6315.00

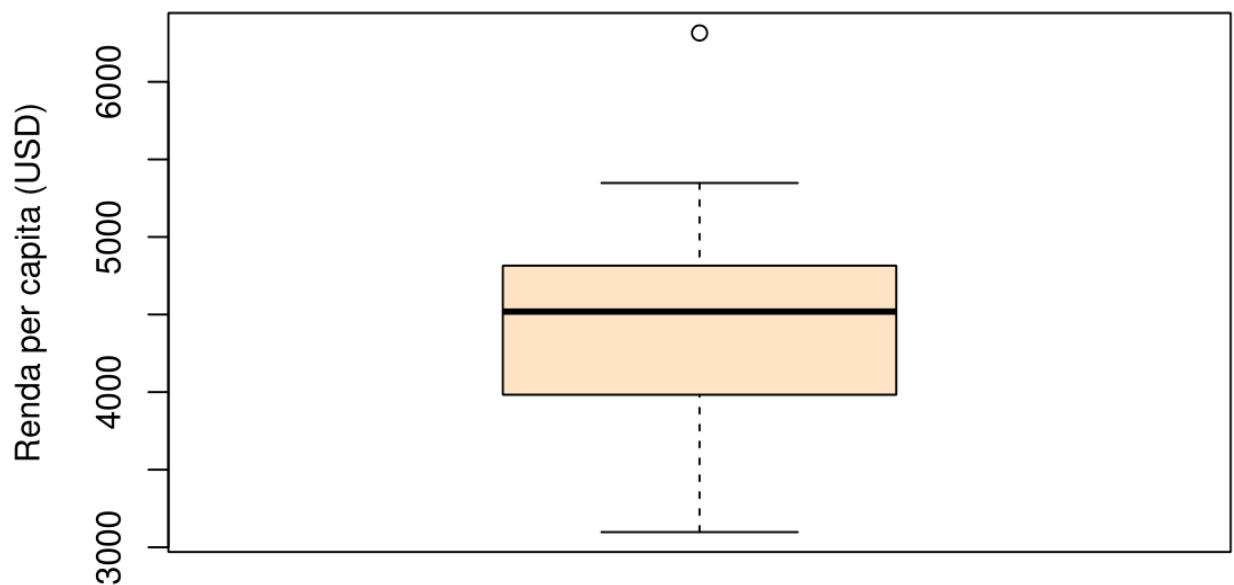


Figura 2. Boxplot relativo à variável quantitativa contínua *renda per capita*.

III. População de analfabetos:

Dos 50 estados avaliados, 18(36%) deles possuem uma taxa de analfabetismo populacional entre 0,5% e 0,78%, sendo que apenas 3(6%) deles têm uma taxa maior que 2,3%. A taxa de analfabetismo nos estados tem média de 0,95%, com taxa mínima de 0,5% e máxima de 2,8%.

Tabela 3.1. Tabela de frequência relativa à variável quantitativa *taxa de analfabetismo*.

Analfabetismo		
Tabela de frequências		
Quantidade	Absoluta	Relativa %
[0.5, 0.7875[18	36
[0.7875, 1.075[8	16
[1.075, 1.3625[7	14
[1.3625, 1.651[5	10
[1.651, 1.9375[5	10
[1.9375, 2.225[4	8
[2.225, 2.5125[2	4
[2.5125, 2.8[1	2

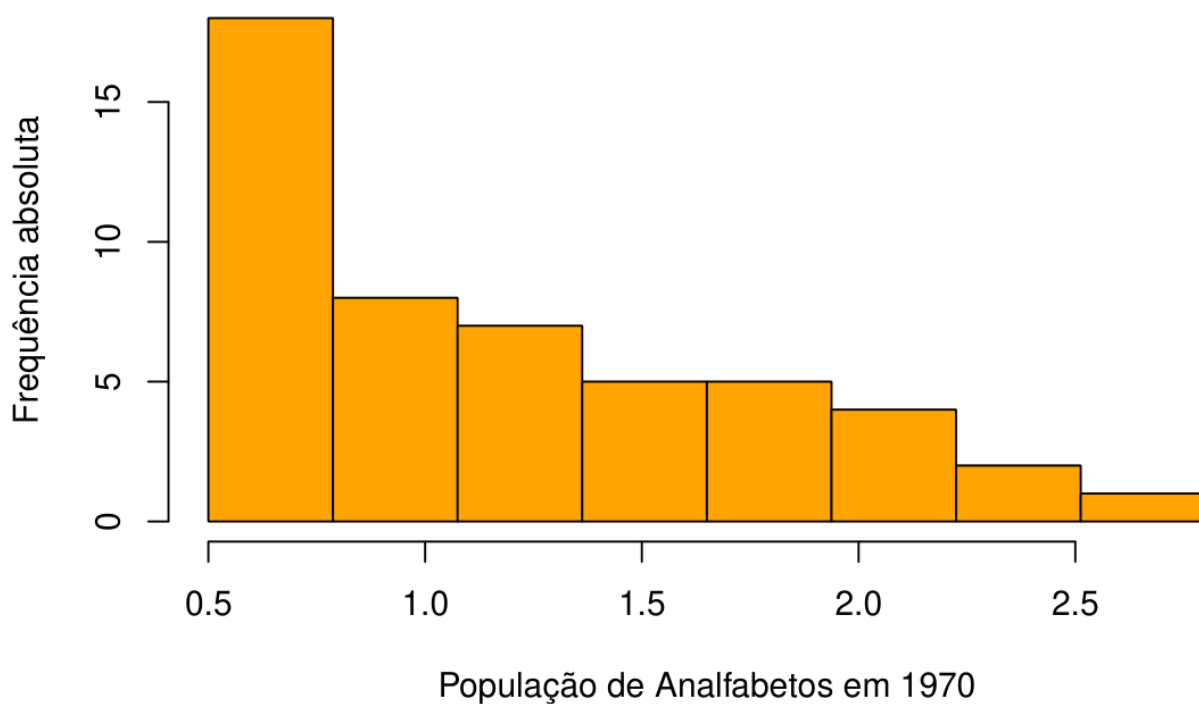


Gráfico 3.1. Histograma relativo à variável quantitativa contínua *taxa de analfabetismo*.

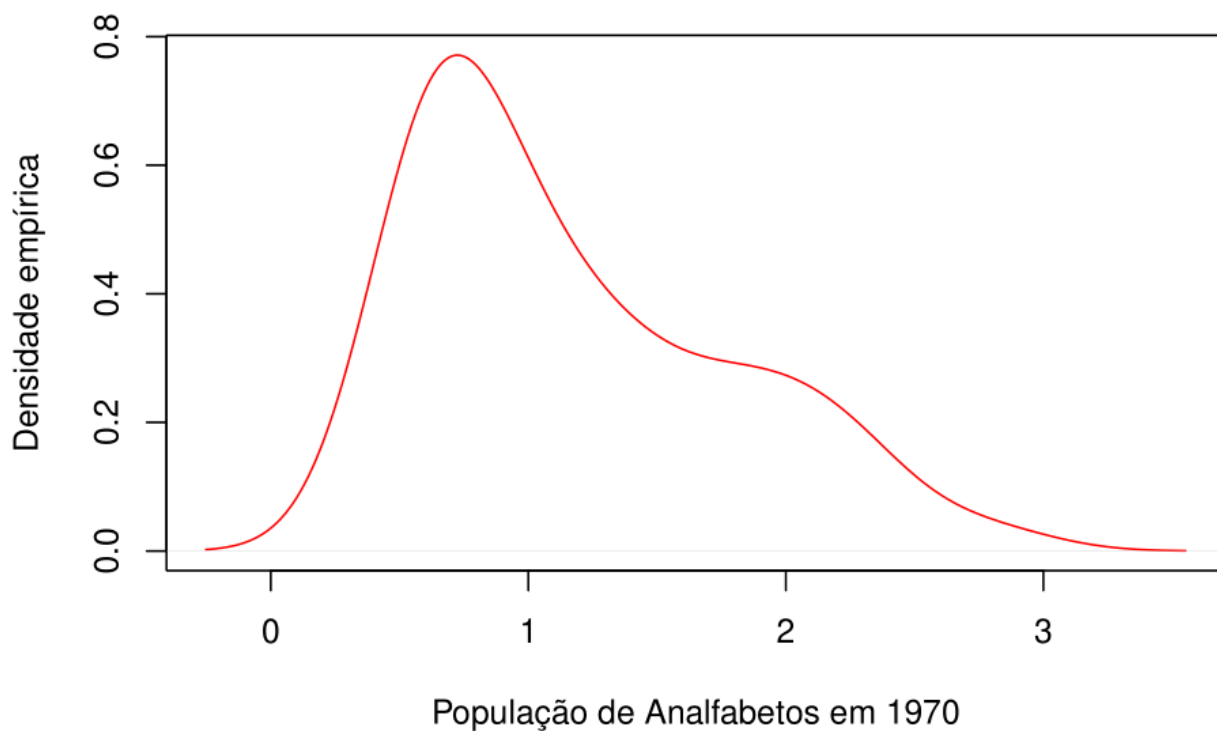


Gráfico 3.2. Gráfico de densidade empírica relativo à variável quantitativa contínua *taxa de analfabetismo*.

Tabela 3.2. Medidas descritivas relativas à variável quantitativa contínua *taxa de analfabetismo*.

Medida	Valor
Min.	0.500
1st Qu.	0.625
Median	0.950
Mean	1.170
3rd Qu.	1.575
Max.	2.800

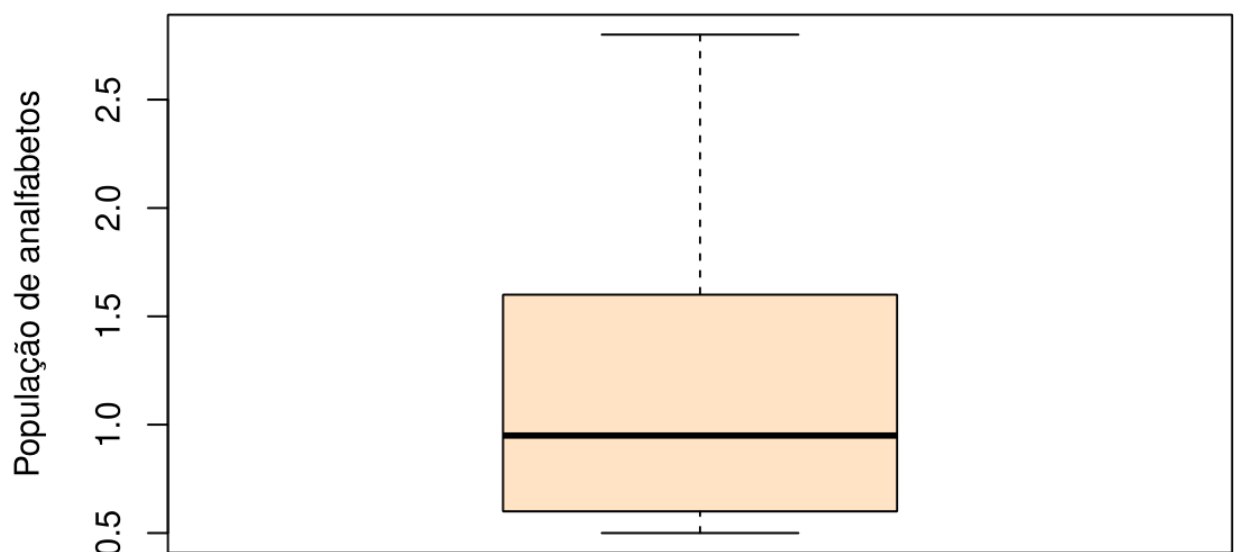


Figura 3. Boxplot relativo à variável quantitativa contínua *taxa de analfabetismo*.

IV. Expectativa de vida:

Dos 50 estados avaliados, 21(42%) deles possuem uma expectativa de vida entre 69.9 e 70.9 anos, sendo que apenas 2(4%) deles têm uma expectativa maior que 72.9 anos. A expectativa de vida nos estados tem média aritmética de 70.6 anos, com mínimo de 67.9 e máximo de 73.9.

Tabela 4.1. Tabela de frequência relativa à variável quantitativa *expectativa de vida*.

Expectativa de Vida		
Tabela de frequências		
Quantidade	Absoluta	Relativa %
[67.96, 68.96[4	8
[68.96, 69.96[5	10
[69.96, 70.96[21	42
[70.96, 71.96[8	16
[71.96, 72.96[10	20
[72.96, 73.96[2	4

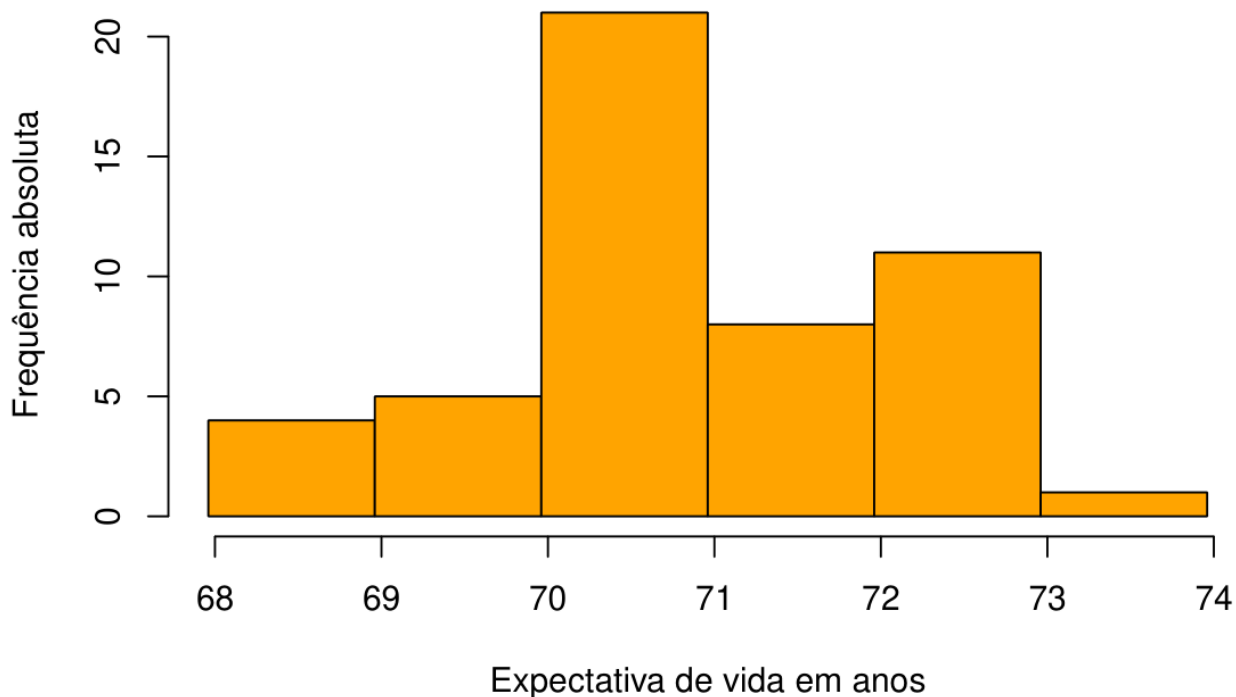


Gráfico 4.1. Histograma relativo à variável quantitativa contínua *expectativa de vida*.



Gráfico 4.2. Gráfico de densidade empírica relativo à variável quantitativa contínua *expectativa de vida*.

Tabela 4.2. Medidas descritivas relativas à variável quantitativa contínua *expectativa de vida*.

Medida	Valor
Min.	67.9600
1st Qu.	70.1175
Median	70.6750
Mean	70.8786
3rd Qu.	71.8925
Max.	73.6000

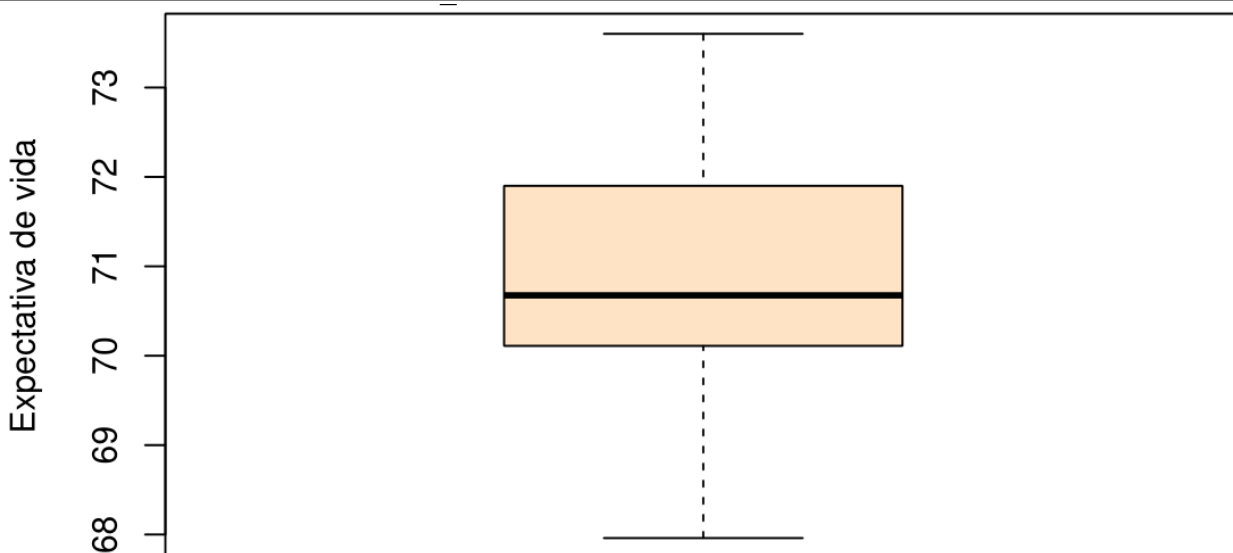


Figura 4. Boxplot relativo à variável quantitativa contínua *expectativa de vida*.

V. Taxa de criminalidade:

Dos 50 estados avaliados, 19(38%) deles possuem uma taxa de criminalidade entre 1.4% e 5.9%, 17(34%) deles possuem taxa entre 5.9% e 10.5%, 13(26%) deles possuem taxa entre 10.5% e 15.1%, sendo que apenas 1(2%) deles têm uma taxa maior que 15.1%. A taxa de criminalidade nos estados tem média aritmética de 6.8%, com mínimo de 1.4% e máximo de 15.1%. A taxa de criminalidade está bastante dispersa nos estados.

Tabela 5.1. Tabela de frequência relativa à variável quantitativa da taxa *de criminalidade*.

Taxa de Criminalidade		
Tabela de frequências		
Quantidade	Absoluta	Relativa %
[1.4, 3.683333[11	22
[3.683333, 5.966667[8	16
[5.966667, 8.25[10	20
[8.25, 10.533333[7	14
[10.533333, 12.816667[11	22
[12.816667, 15.1]	2	4
[15.1, 17.3[1	2

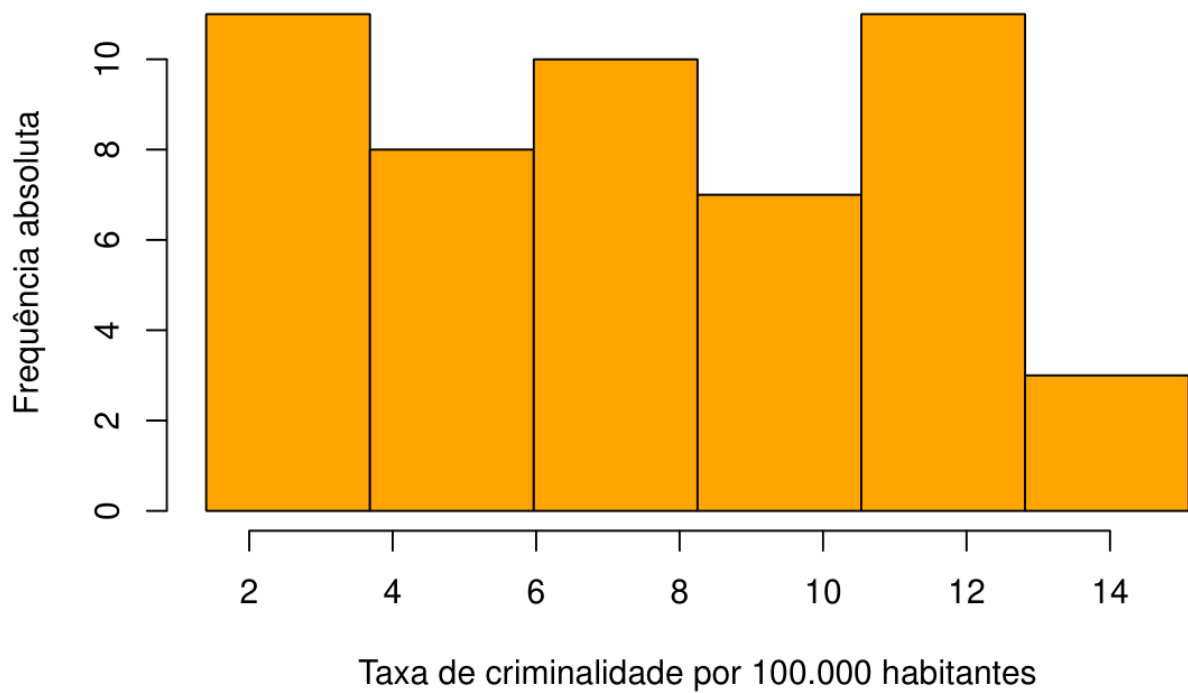


Gráfico 5.1. Histograma relativo à variável quantitativa contínua *taxa de criminalidade*.

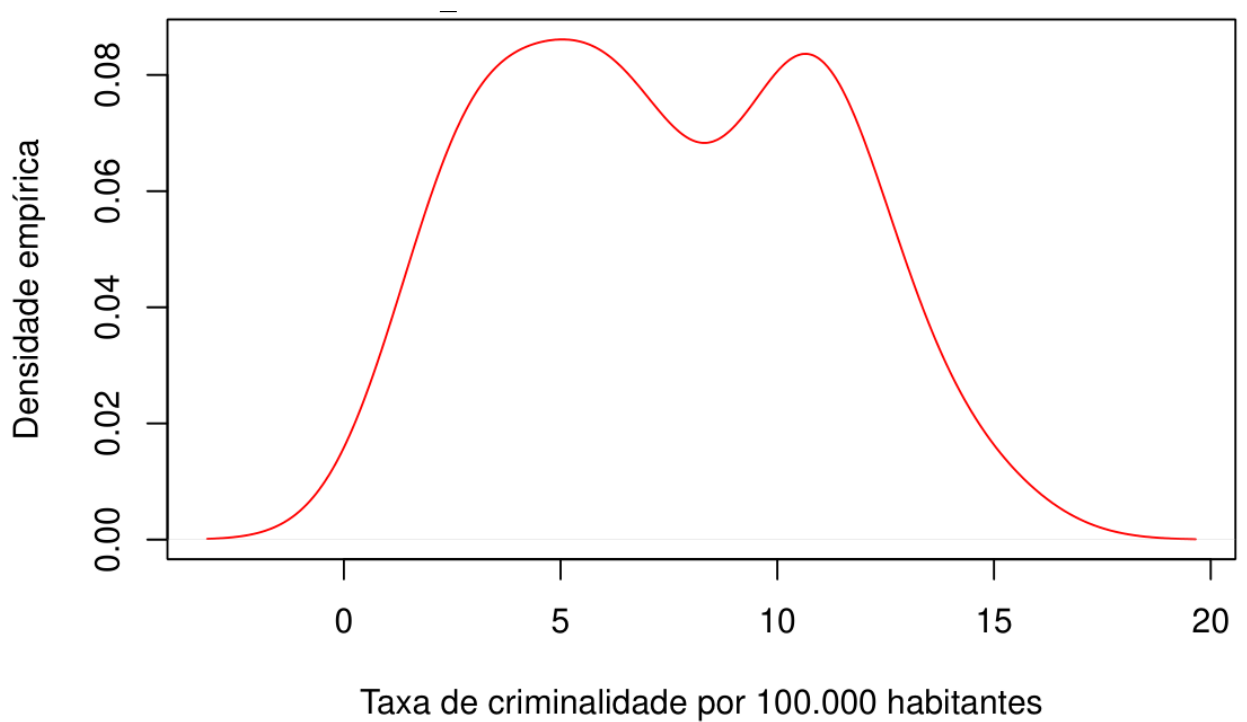
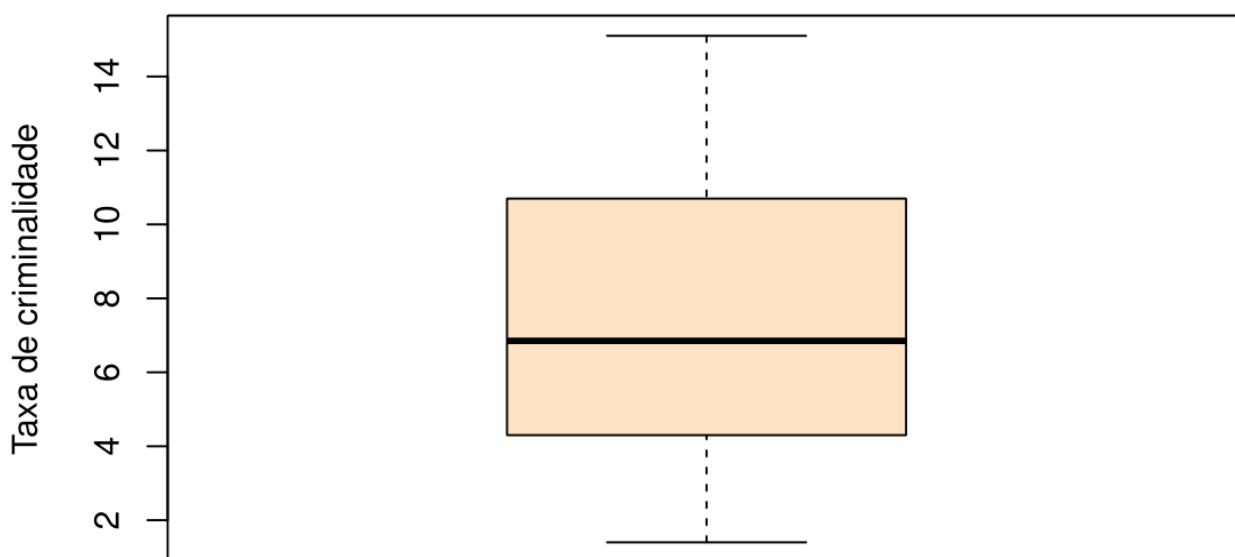


Gráfico 5.2. Gráfico de densidade empírica relativo à variável quantitativa contínua *taxa de criminalidade*.

Tabela 5.2. Medidas descritivas relativas à variável quantitativa contínua taxa *de criminalidade*.

Medida	Valor
Min.	1.400
1st Qu.	4.350
Median	6.850
Mean	7.378
3rd Qu.	10.675
Max.	15.100

**Figura 5.** Boxplot relativo à variável quantitativa contínua *expectativa de vida*.

VI. Estudantes que concluem o ensino médio:

Entre os 50 estados avaliados, cerca de 14(28%) estados apresentam uma população com um índice de conclusão do segundo grau abaixo de aproximadamente 50%, enquanto 28(56%) estados apresentam um índice aproximado de uma intervalo de 50% a 60%, por fim, 8(16%) estados apresentam índices aproximados superiores a 60%. O índice médio se aproxima de 53%, enquanto o estado com menor índice apresentou uma taxa de 37.80% o de maior apresentou uma de 67.30%. Veja, Tabelas 6.1 e 6.2 e Gráficos 6, 6.1 e 6.2.

Tabela 6.1. Tabela de frequência relativa à variável quantitativa *taxa de conclusão do segundo grau*.

Taxa de conclusão do segundo grau

Tabela de frequências

Quantidade	Absoluta	Relativa %
[37.8, 40.75[5	10
[40.75, 43.7[5	10
[43.7, 46.65[1	2
[46.65, 49.6[3	6
[49.6, 52.55[5	10
[52.55, 55.5[11	22
[55.5, 58.45[5	10
[58.45, 61.4[7	14
[61.4, 64.35[5	10
[64.35, 67.3[3	6

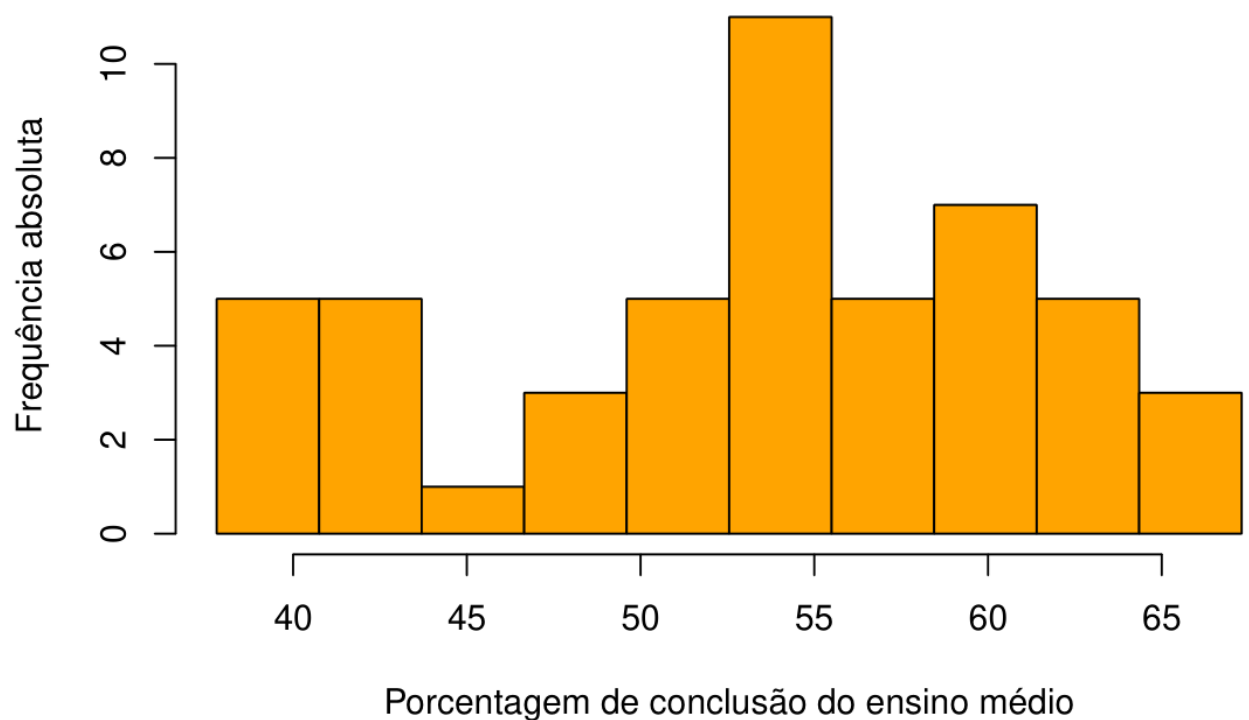


Gráfico 6.1. Histograma relativo à variável quantitativa contínua *taxa de conclusão do segundo grau*.

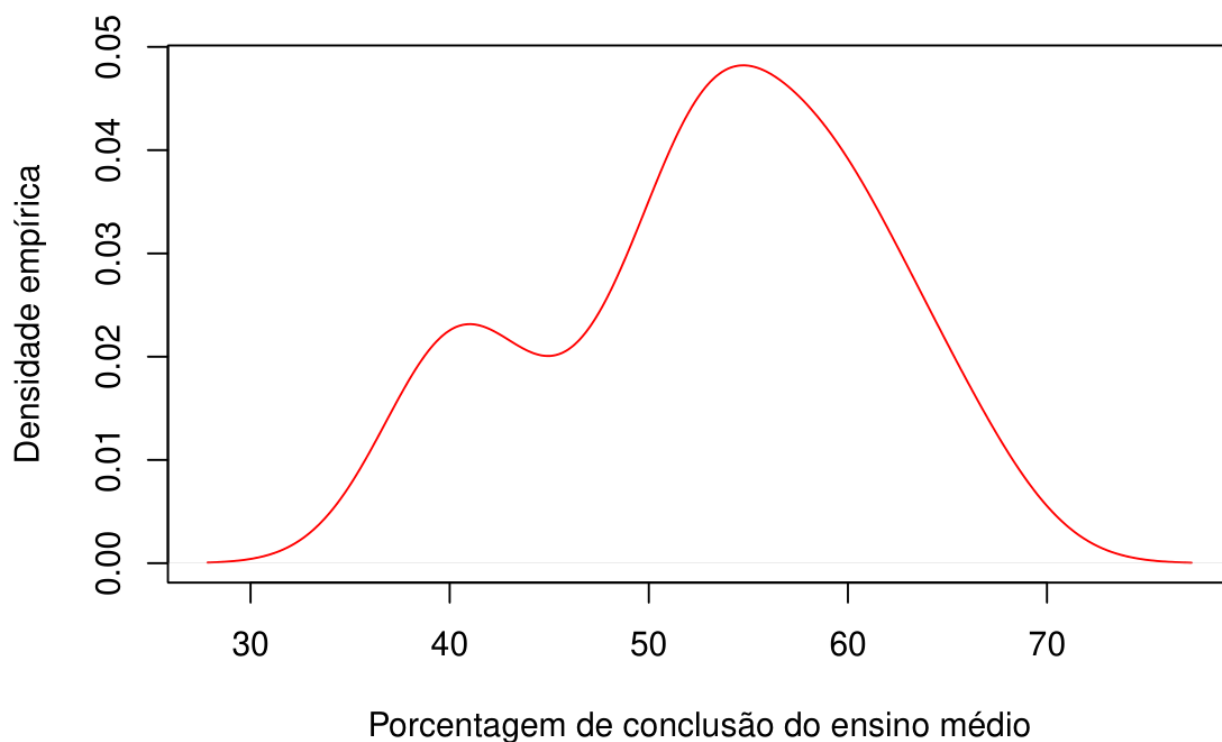


Gráfico 6.2. Gráfico de densidade empírica relativo à variável quantitativa contínua *taxa de conclusão do segundo grau*.

Tabela 6.2. Medidas descritivas relativas à variável quantitativa contínua *taxa de conclusão do segundo grau*.

	Medida	Valor
1	Min.	37.800
2	1st Qu.	48.050
3	Median	53.250
4	Mean	53.108
5	3rd Qu.	59.150
6	Max.	67.300



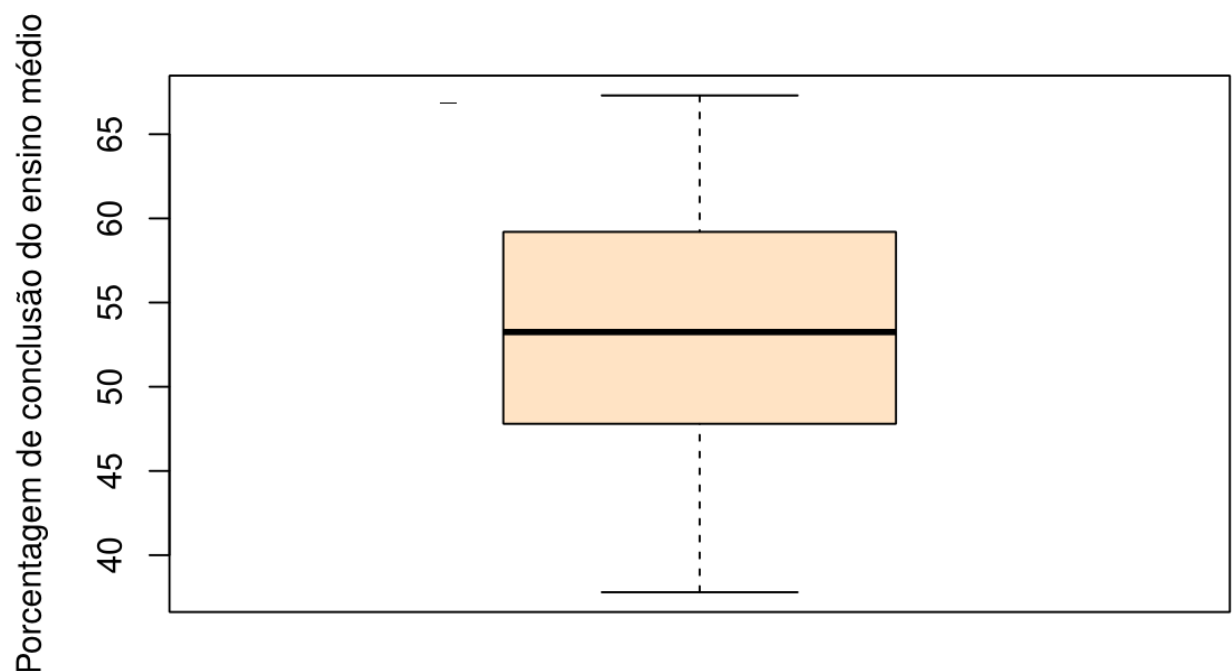


Figura 6. Boxplot relativo à variável quantitativa contínua *taxa de conclusão do segundo grau*.

VII. Número de dias do ano com temperatura abaixo de 0:

Os dados relativos aos dias são bastante particulares de cada estado e espaçados, tal fato se deve a relação entre localização do estado e temperatura. Dessa forma, percebe-se que os estados mais ao norte (maior latitude) apresentam uma quantidade maior de dias com temperatura abaixo de 0 graus Celsius que os estados localizados mais ao sul (menor latitude). O menor número de dias registrado foi 0 relativo ao Havaí (território anexado, fruto da ação imperialista dos Estados Unidos, que se localiza em um arquipélago no Oceano Pacífico, localizado fora do território da América do Norte), a média é de 114 dias aproximadamente e o máximo registrado foi de 188 dias.

Tabela 6.2 Tabela de frequência variável *número de dias com temperatura abaixo de 0°C.*

Dias com temperatura abaixo de 0 graus Celsius.

Tabela de frequências

Estados	Absoluta	Relativa %
Alabama	20	0.38
Alaska	152	2.91
Arizona	15	0.29
Arkansas	65	1.24
California	20	0.38
Colorado	166	3.18
Connecticut	139	2.66
Delaware	103	1.97
Florida	11	0.21
Georgia	60	1.15
Hawaii	0	0
Idaho	126	2.41
Illinois	127	2.43
Indiana	122	2.34
Iowa	140	2.68
Kansas	114	2.18
Kentucky	95	1.82
Louisiana	12	0.23
Maine	161	3.08
Maryland	101	1.93
Massachusetts	103	1.97
Michigan	125	2.39
Minnesota	160	3.06
Mississippi	50	0.96

Missouri	108	2.07
Montana	155	2.97
Nebraska	139	2.66
Nevada	188	3.6
New-Hampshire	174	3.33
New-Jersey	115	2.2
New-Mexico	120	2.3
New-York	82	1.57
North-Carolina	80	1.53
North-Dakota	186	3.56
Ohio	124	2.37
Oklahoma	82	1.57
Oregon	44	0.84
Pennsylvania	126	2.41
Rhode-Island	127	2.43
South-Carolina	65	1.24
South-Dakota	172	3.29
Tennessee	70	1.34
Texas	35	0.67
Utah	137	2.62
Vermont	168	3.22
Virginia	85	1.63
Washington	32	0.61
West-Virginia	100	1.91
Wisconsin	149	2.85
Wyoming	173	3.31

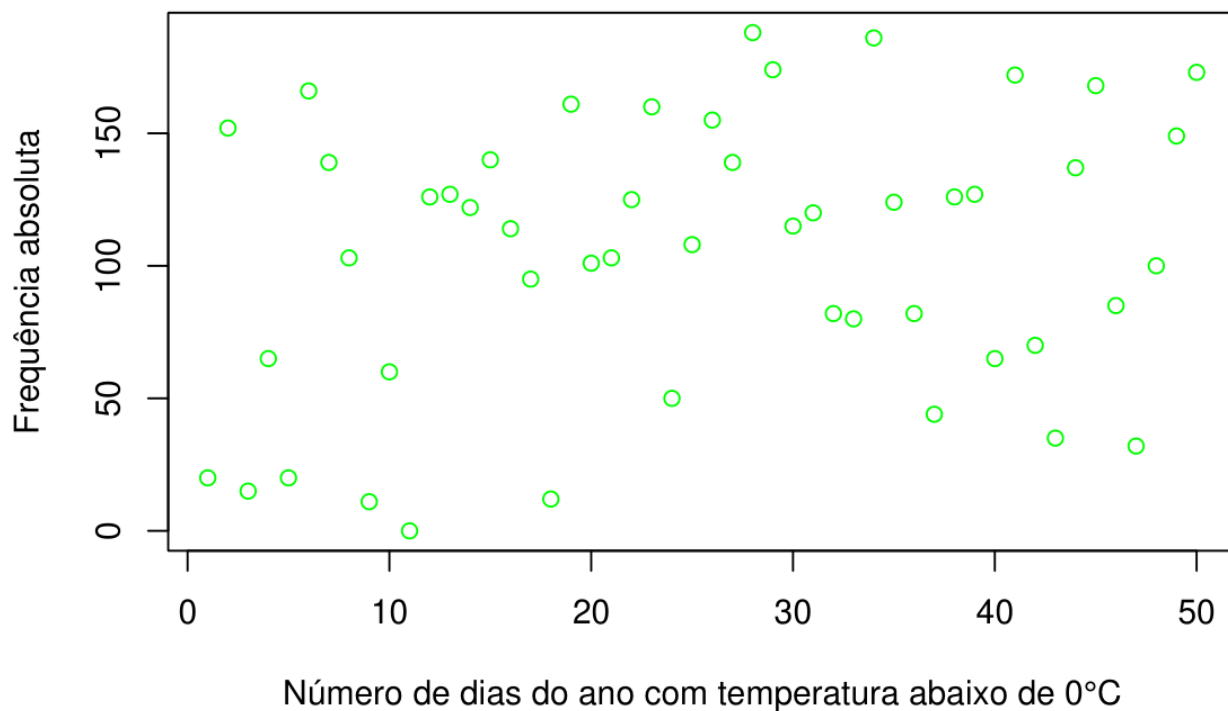


Gráfico 7.1. Gráfico relativo à variável quantitativa discreta *número de dias do ano com temperatura abaixo de 0°C*.

Tabela 7.2. Medidas descritivas relativas à variável quantitativa discreta *número de dias do ano com temperatura abaixo de 0°C*.

Medida	Valor
Min.	0.00
1st Qu.	66.25
Median	114.50
Mean	104.46
3rd Qu.	139.75
Max.	188.00

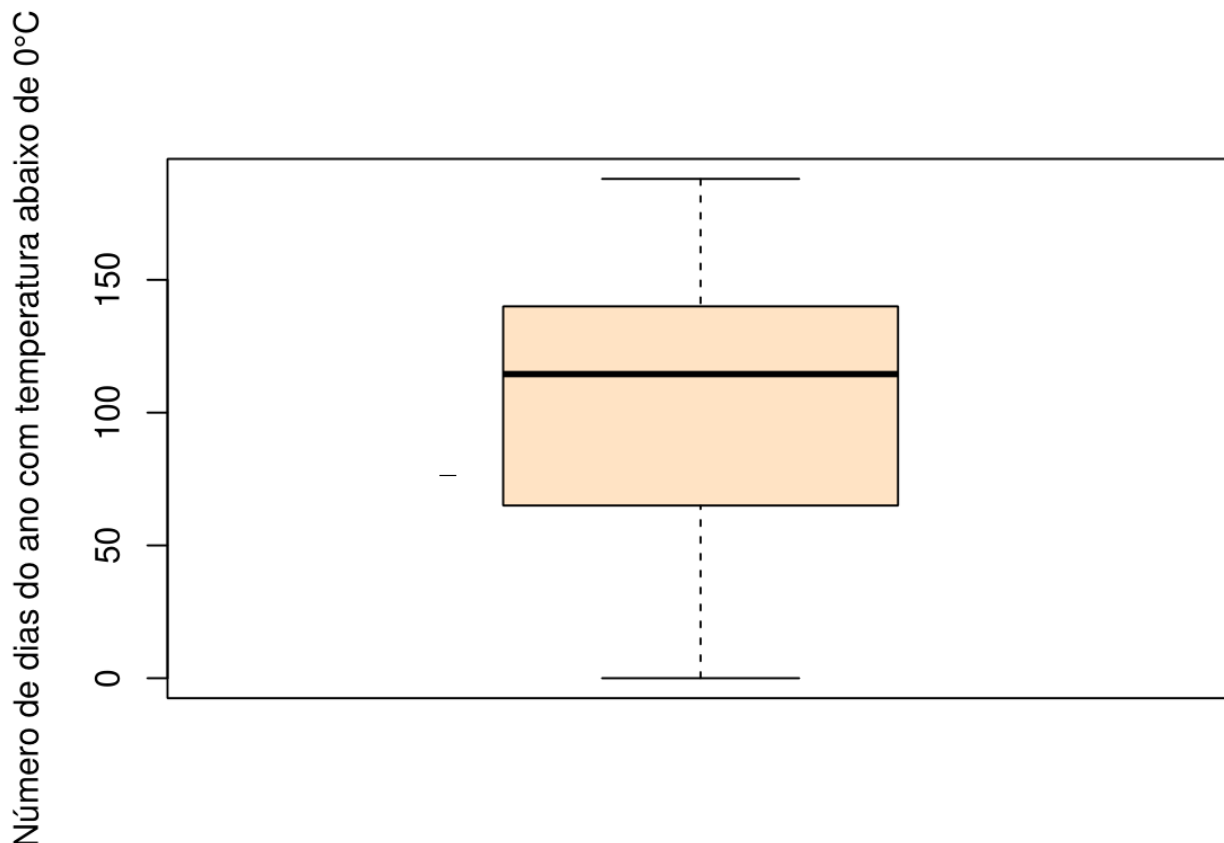


Figura 7. Boxplot relativo à variável quantitativa discreta *número de dias do ano com temperatura abaixo de 0°C*.

VIII. Área do estado:

Em relação às áreas dos estados, em milhas, 30(60%) estados possuem uma área dentro do intervalo de 0 a 60000 milhas, 15(30%) estados possuem áreas entre 60.000 e 120.000 milhas, 2 (4%) estados com áreas entre 120.000 e 180.000 milhas, 2(4%) estados com áreas entre 240.000 e 300.000 milhas e 1(2%) estado com área entre 540.000 e 600.000 milhas. O menor estado possui uma área de cerca de 1049 milhas, o maior estado possui área de 655.432 milhas e a média nacional foi de 54.277 milhas.

Tabela 8.1. Tabela de frequência relativa à variável quantitativa *área em milhas*.

Taxa de área em milhas		
Tabela de frequências		
Quantidade	Absoluta	Relativa %
[0, 60000[30	60
[60000, 120000[15	30
[120000, 180000[2	4
[180000, 240000[0	0
[240000, 300000[2	4
[300000, 360000[0	0
[360000, 420000[0	0
[420000, 480000[0	0
[480000, 540000[0	0
[540000, 600000[1	2



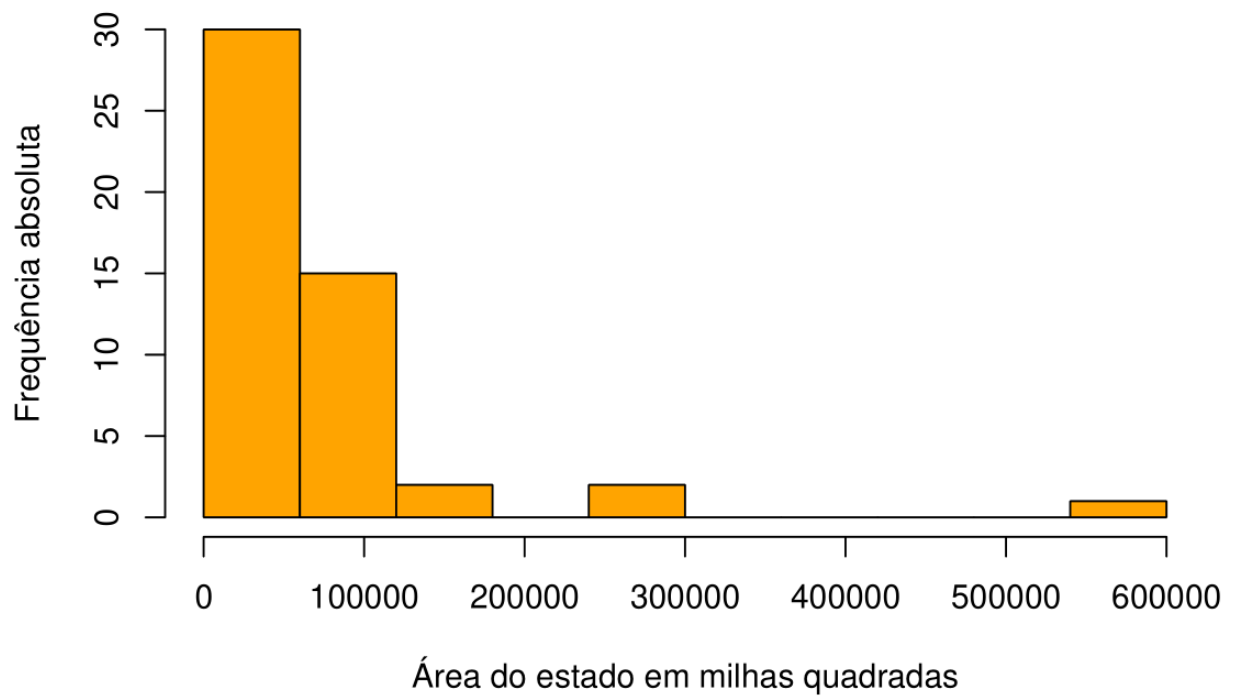


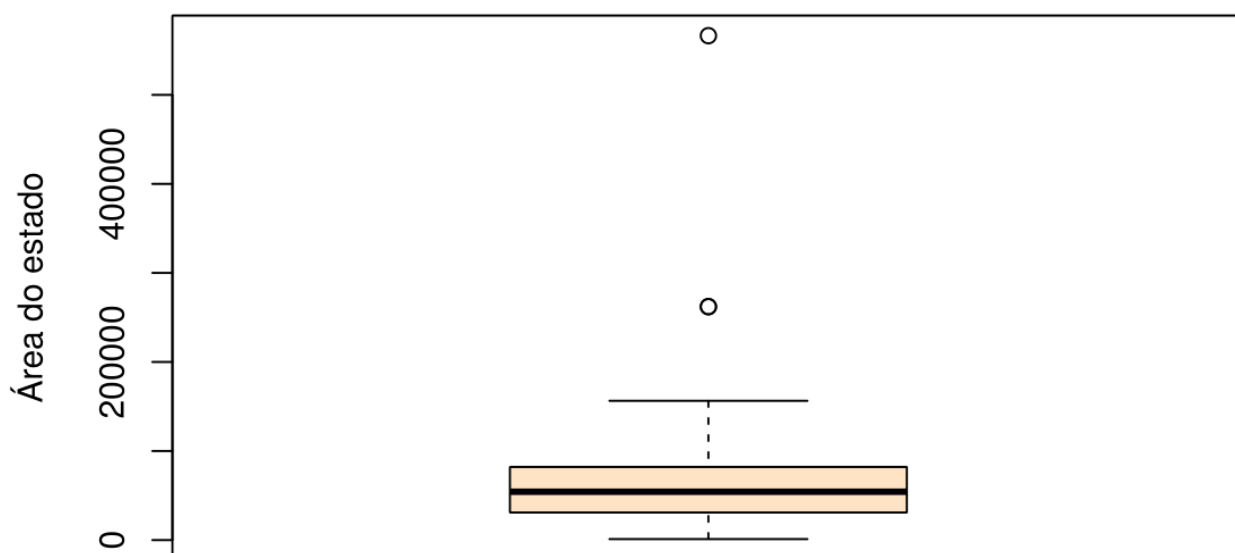
Gráfico 8.1. Histograma relativo à variável quantitativa contínua *área em milhas*.



Gráfico 8.2. Gráfico de densidade empírica relativo à variável quantitativa contínua *área em milhas*.

Tabela 8.2. Medidas descritivas relativas à variável quantitativa contínua *área em milhas*.

Medida	Valor
Min.	1049.00
1st Qu.	32214.25
Median	54277.00
Mean	72862.20
3rd Qu.	82018.75
Max.	566432.00

**Figura 8.** Boxplot relativo à variável quantitativa contínua *área em milhas*.

IX. Densidade populacional:

Em relação a densidade populacional têm-se que 45(90%) estados possuem uma densidade entre 0 e 0.44, 3(6%) estados possuem um índice entre 0.44 e 0.89, 1(2%) estado entre 0.89 e 1.34 e 1 estado com índice entre 2.69 e 3.13, o cálculo é feito com base em: $densidade = população / área$. O estado com menor densidade possui um índice de 0.000064, o estado maior densidade tem índice de 2.68 e a média nacional foi de 0.06.

Tabela 9.1. Tabela de frequência relativa à variável quantitativa *densidade populacional*.

Densidade		
Tabela de frequências		
Quantidade	Absoluta	Relativa %
[0.0000000, 0.4479472[45	90
[0.4479472, 0.89525[3	6
[0.89525, 1.3425528[1	2
[1.3425528, 1.7898556[0	0
[1.7898556, 2.2371584[0	0
[2.2371584, 2.6844614[0	0
[2.6844614, 3.1319168[1	2

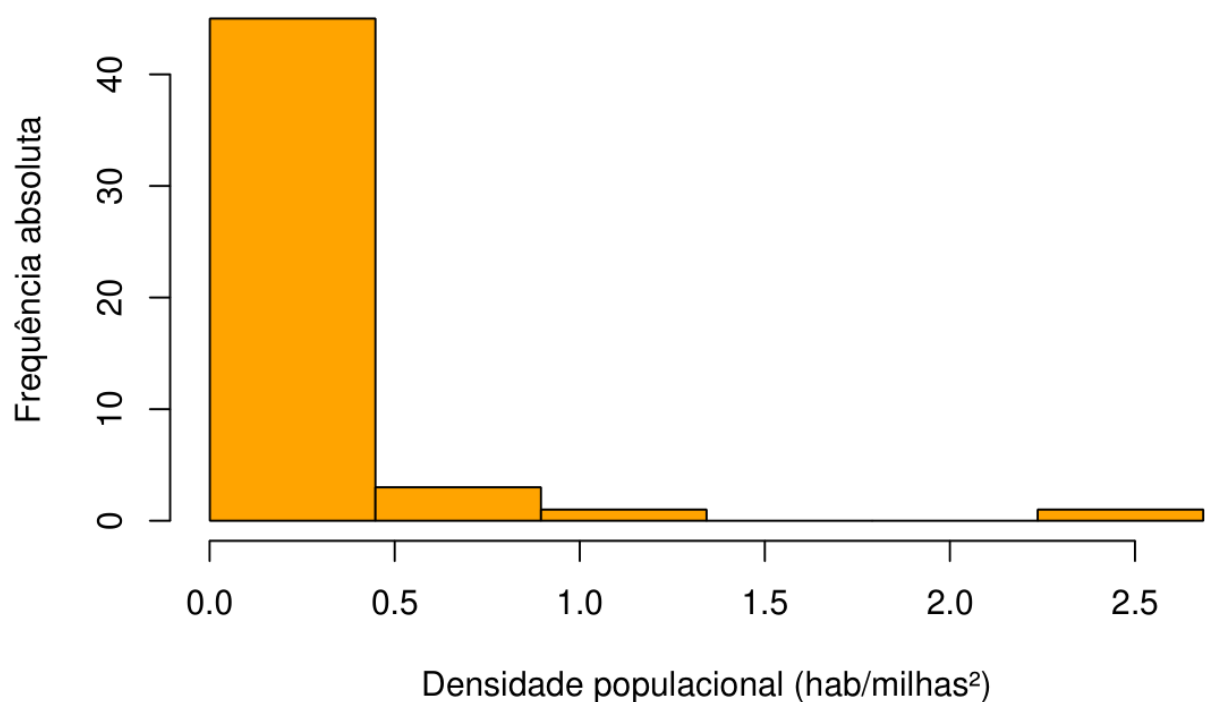


Gráfico 9.1. Histograma relativo à variável quantitativa contínua *densidade populacional*.

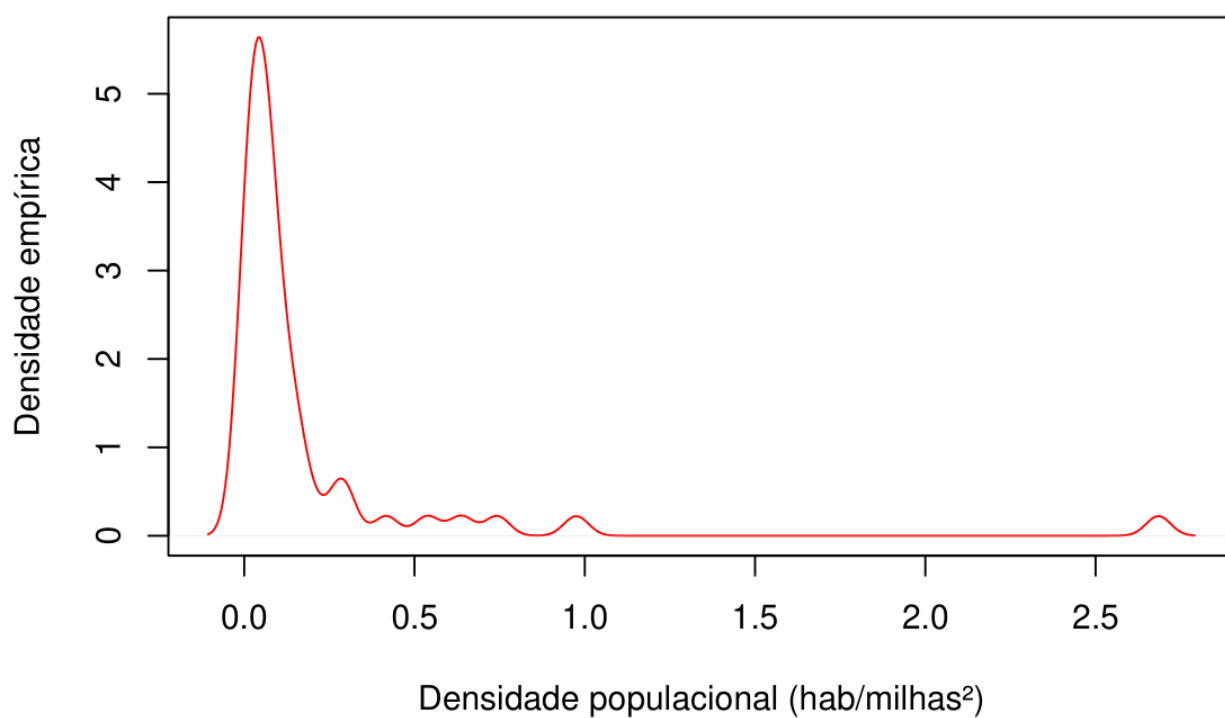


Gráfico 9.2. Gráfico de densidade empírica relativo à variável quantitativa contínua *densidade populacional*.

Medida	Valor
Min.	0.0006443845
1st Qu.	0.0277300661
Median	0.0690245307
Mean	0.1879501822
3rd Qu.	0.1442827827
Max.	2.6844613918

Tabela 9.2. Medidas descritivas relativas à variável quantitativa contínua *densidade populacional*.

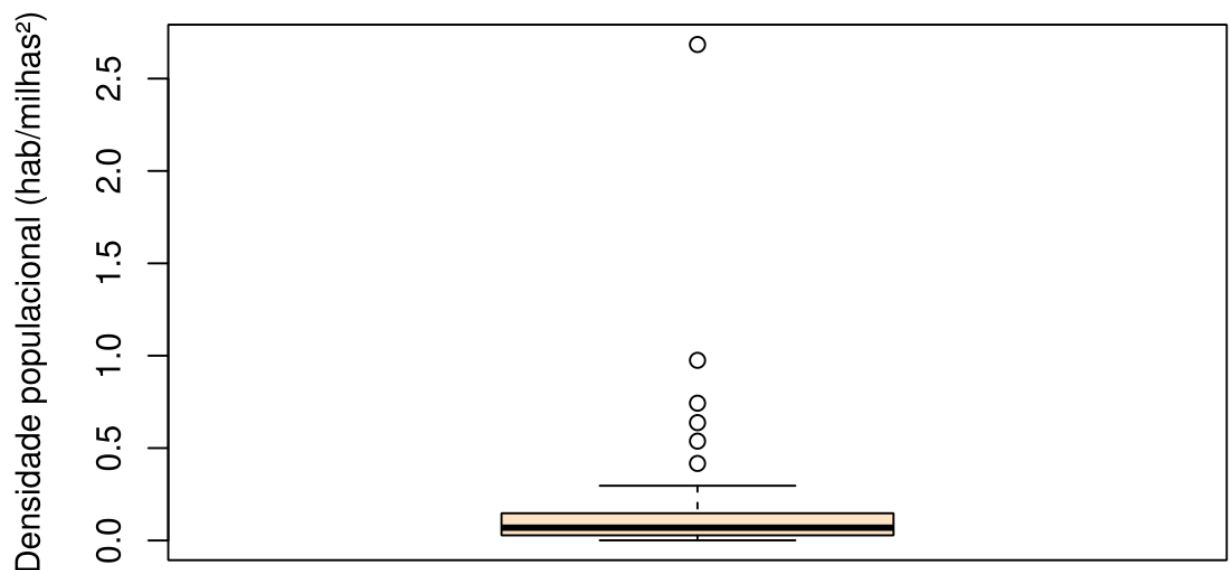


Figura 9. Boxplot relativo à variável quantitativa contínua *densidade populacional*.

3 ANÁLISE DESCRITIVA BIDIMENSIONAL

1. Renda per capita:

Em geral, a renda per capita apresentou uma correlação positiva baixa (0,34) com a expectativa de vida. Como pode ser constatado no gráfico 10.1, nas faixas mais baixas de renda a correlação parece ser um pouco mais significativa, porém ainda insuficiente para inferir conclusões precisas. Os dados se tornam mais difusos ainda a partir da faixa dos \$4000 per capita. Dessa forma, pelo menos na amostra coletada, a renda per capita não apresentou uma correlação significativa com a expectativa de vida.

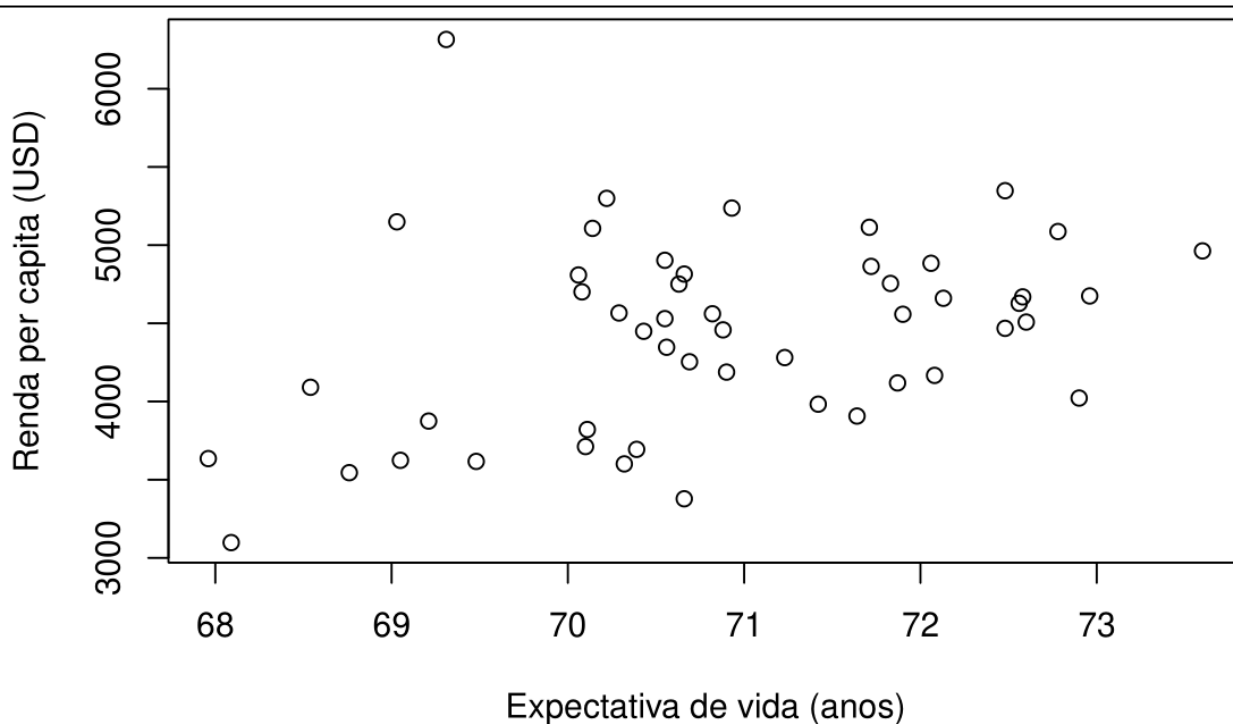


Gráfico 10.1. Gráfico de correlação entre expectativa de vida e renda per capita.

2. Taxa de analfabetismo:

A taxa de analfabetismo apresenta uma correlação negativa (-0,59) com a expectativa de vida. Ademais, vale destacar que na amostra coletada nos extremos do gráfico parece haver uma clara correlação entre a taxa de analfabetismo e a expectativa de vida. Apesar das tendências constatadas nos extremos, no resto do gráfico constata-se uma correlação fraca entre as duas variáveis e uma difusão dos dados. Por fim, é sabido que a taxa de analfabetismo pode ser um indicativo para outros fatores que possam impactar de forma mais direta na expectativa de vida, porém na amostra coletada ela não apresenta uma correlação forte o bastante para se afirmar uma implicação direta.

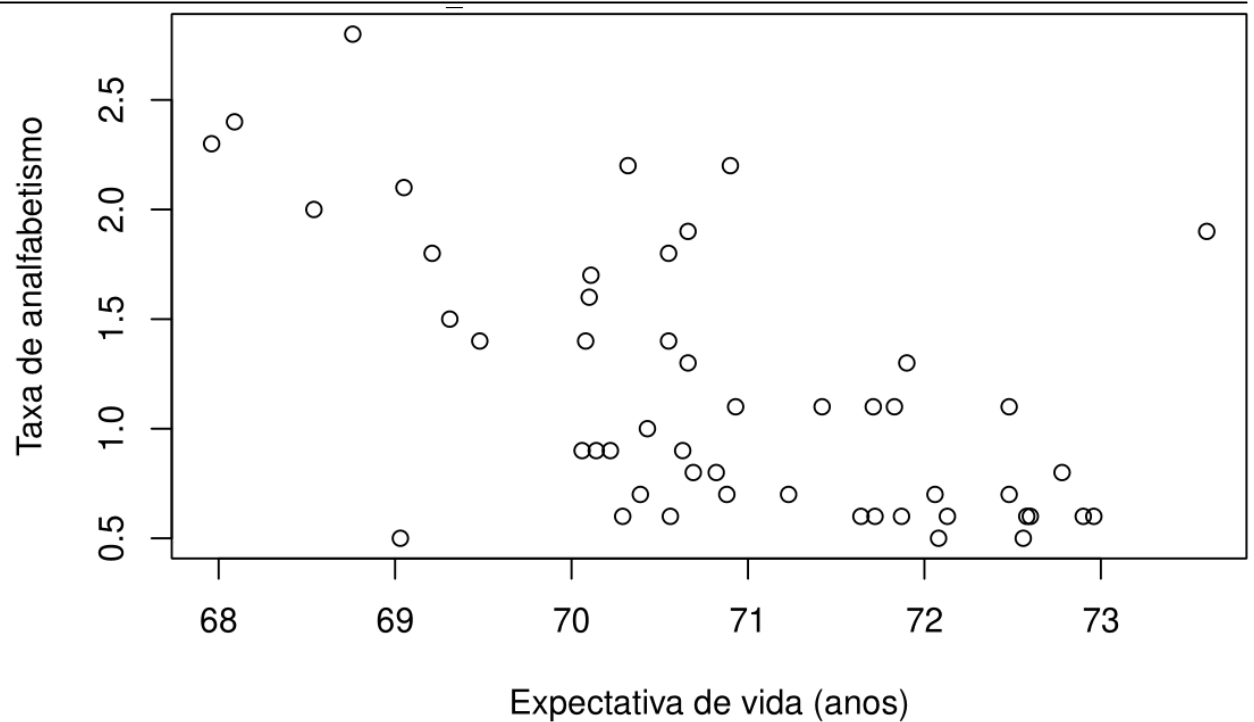


Gráfico 10.2. Gráfico de correlação entre expectativa de vida e taxa de analfabetismo.

3. Taxa de criminalidade:

Dentre todas as variáveis analisadas, a taxa de criminalidade estabelece a correlação negativa mais forte (-0,78) com a expectativa de vida. Assim sendo, é possível constatar através do gráfico 10.3 uma tendência indicando que, em geral, uma taxa de criminalidade mais baixa está associada a uma expectativa de vida mais alta. Vale ressaltar que, apesar de ser possível constatar uma tendência, na amostra coletada alguns estados apresentam um desvio significativo dessa tendência.

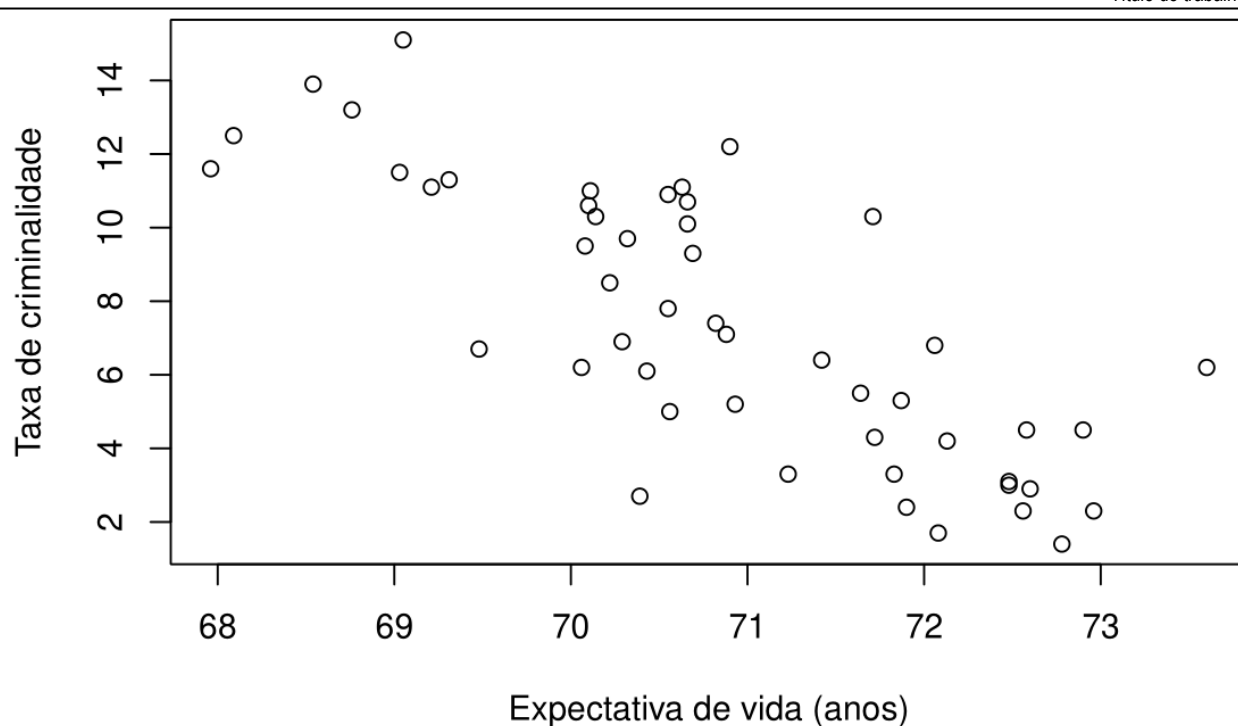


Gráfico 10.3. Gráfico de correlação entre expectativa de vida e taxa de criminalidade.

4. Porcentagem de estudantes que concluem o ensino médio:

A porcentagem de estudantes que concluem o ensino médio apresenta uma correlação positiva de 0,58 com a expectativa de vida. Ao observar o gráfico 10.4 é possível constatar uma dispersão dos dados e, em geral, uma baixa taxa de correlação. Apesar dessa dispersão geral, nota-se que nas faixas mais altas do gráfico (tanto de porcentagem de conclusão do ensino médio quanto de expectativa de vida) parece haver uma correlação mais forte entre as duas variáveis em comparação às faixas mais baixas do gráfico, nas quais não constata-se correlação nenhuma entre ambas as variáveis.

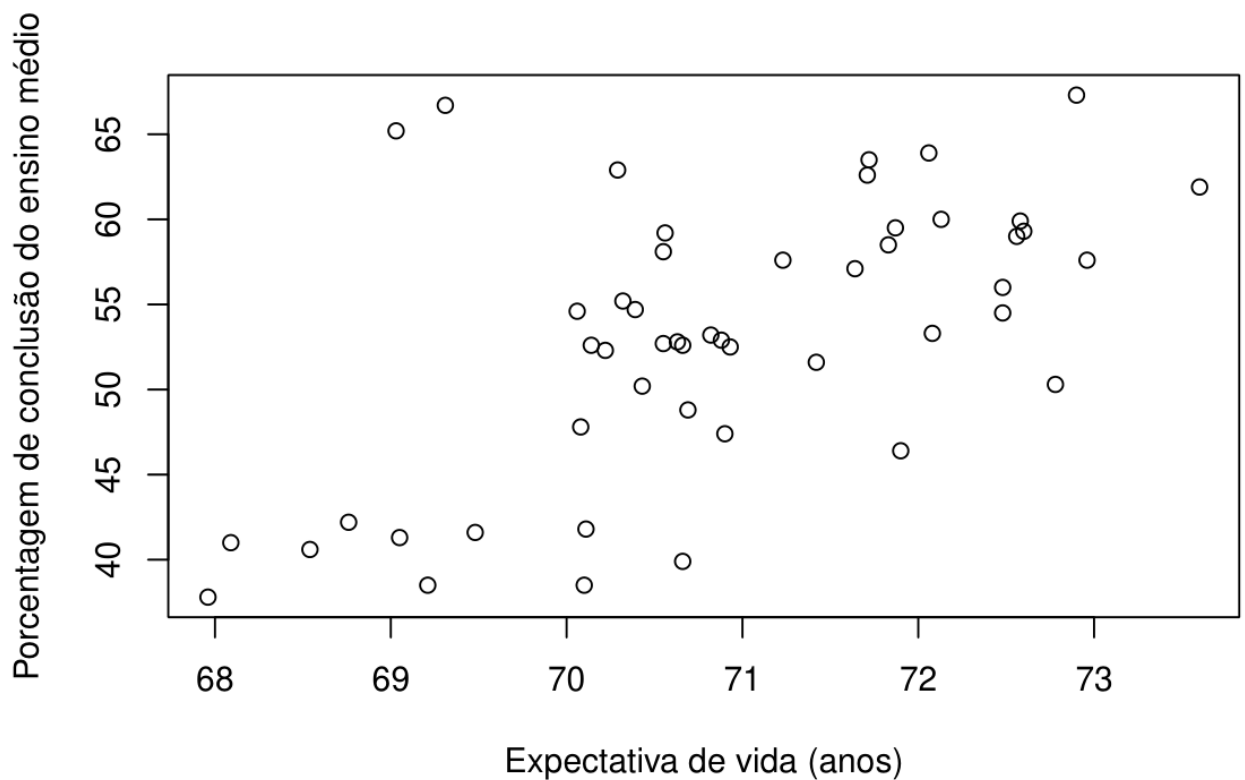


Gráfico 10.4. Gráfico de correlação entre expectativa de vida e porcentagem de conclusão do ensino médio.

5. Dias do ano com temperatura abaixo de 0°C:

O número de dias do ano com temperatura abaixo de zero estabelece a correlação mais fraca com a expectativa de vida, com uma taxa de correlação empírica de 0,26. Através do gráfico é possível constatar uma dispersão alta dos dados, decorrente da baixa correlação entre as duas variáveis.

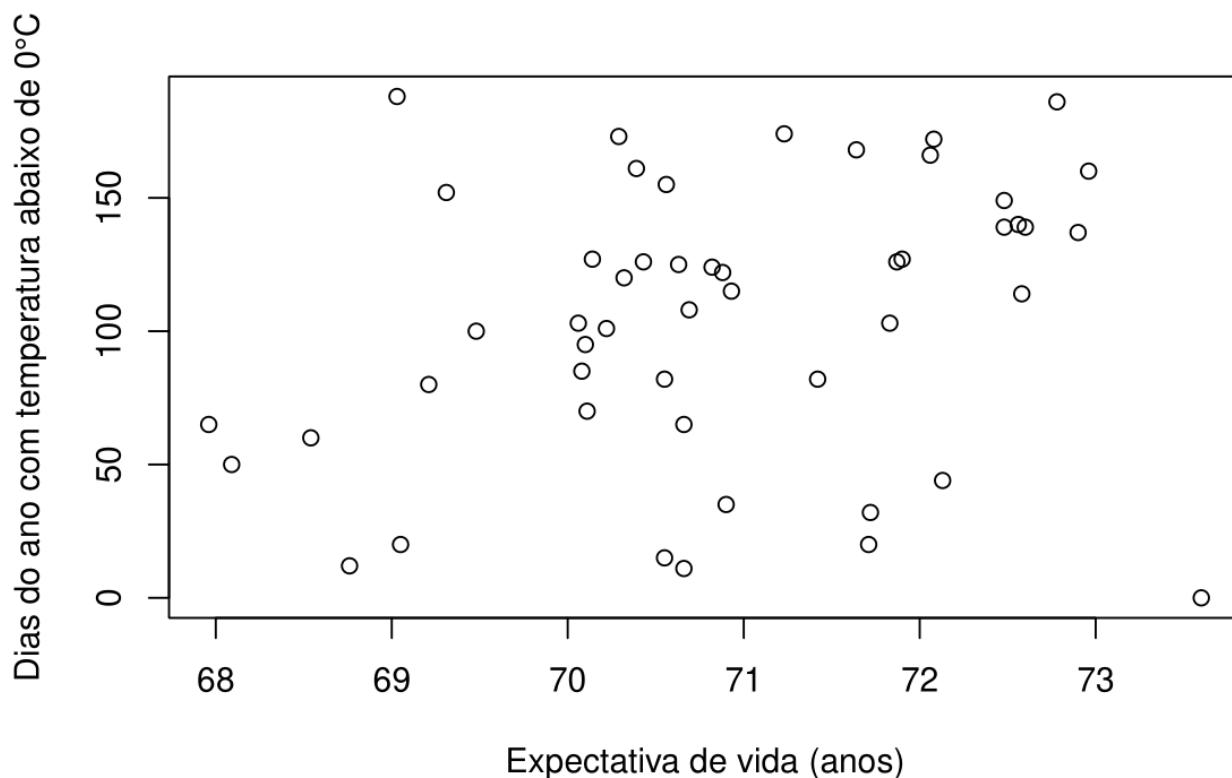


Gráfico 10.5. Gráfico de correlação entre expectativa de vida e dias do ano com temperatura abaixo de 0°C.

6. Densidade populacional:

A densidade populacional apresentou a correlação negativa mais baixa com a expectativa de vida, com uma taxa de correlação empírica de -0,26. Além disso, pode ser constatado através do gráfico 10.6 a dispersão nos dados e a concentração dos dados da densidade populacional na faixa dos valores mais baixos do eixo das ordenadas.

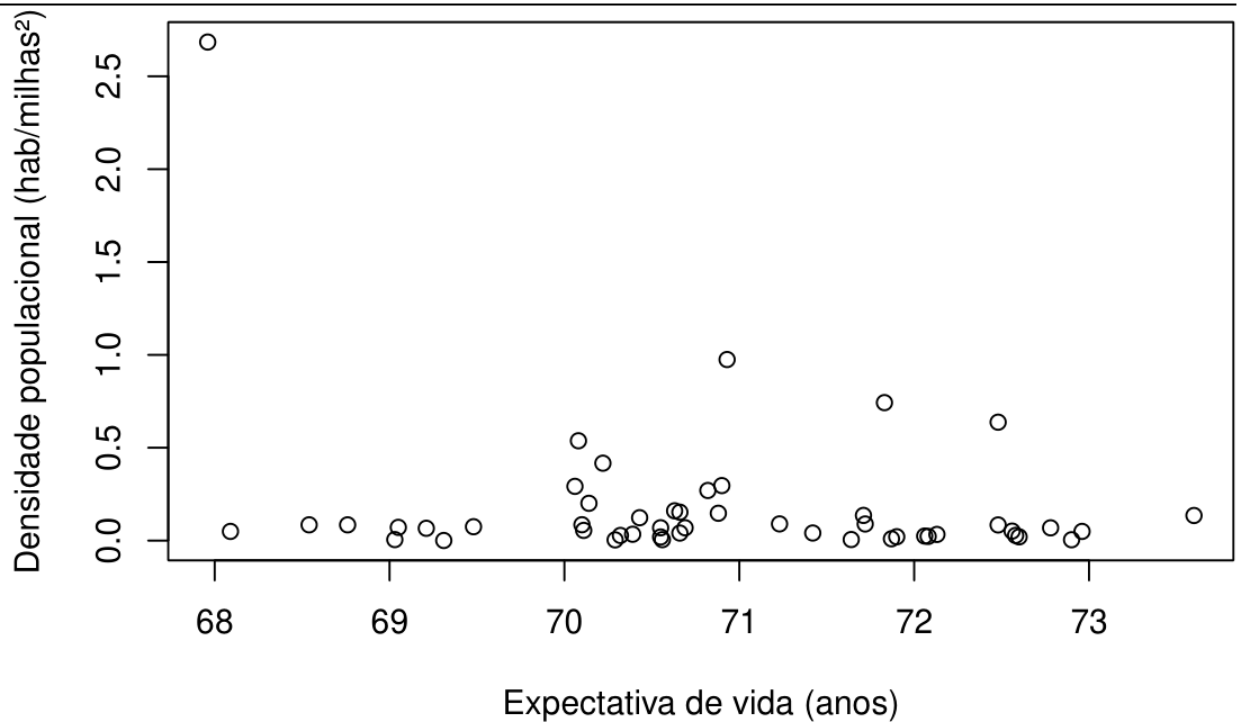


Gráfico 10.6. Gráfico de correlação entre expectativa de vida e densidade populacional.

4 ANÁLISE DE REGRESSÃO LINEAR SIMPLES

Nesta seção, vamos conduzir algumas análises de regressão linear simples para determinar quais variáveis são importantes para explicar a variabilidade da **expectativa de vida**. Assumimos um nível de significância de 5% para os testes de hipóteses conduzidos para este fim. Com o intuito de verificar se a regressão é linear, realizou-se para cada tipo de dado dois testes de hipótese relativos às variáveis β_0 , que representa o valor esperado da expectativa de vida média (variável resposta) quando a variável explicativa é zero e β_1 , que representa a mudança na média dessa expectativa de vida quando aumenta-se uma unidade no valor da variável explicativa.

1. Expectativa de vida e renda per capita:

Os resultados dos testes de hipótese para as variáveis β_0 e β_1 são apresentados na Figura 10.1.

```

Residuals:
    Min       1Q   Median       3Q      Max
-2.96547 -0.76381 -0.03428  0.92876  2.32951

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 6.758e+01  1.328e+00  50.906  <2e-16 ***
percap      7.433e-04  2.965e-04   2.507   0.0156 *
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Figura 10.1. Resultado dos testes de hipótese de regressão linear entre expectativa de vida e a renda per capita.

Analisando os testes de hipótese entre a expectativa de vida e a renda per capita (Figura 10.1), observamos que é razoável assumir que existe uma relação linear entre as variáveis em questão com um nível de significância de 5%.

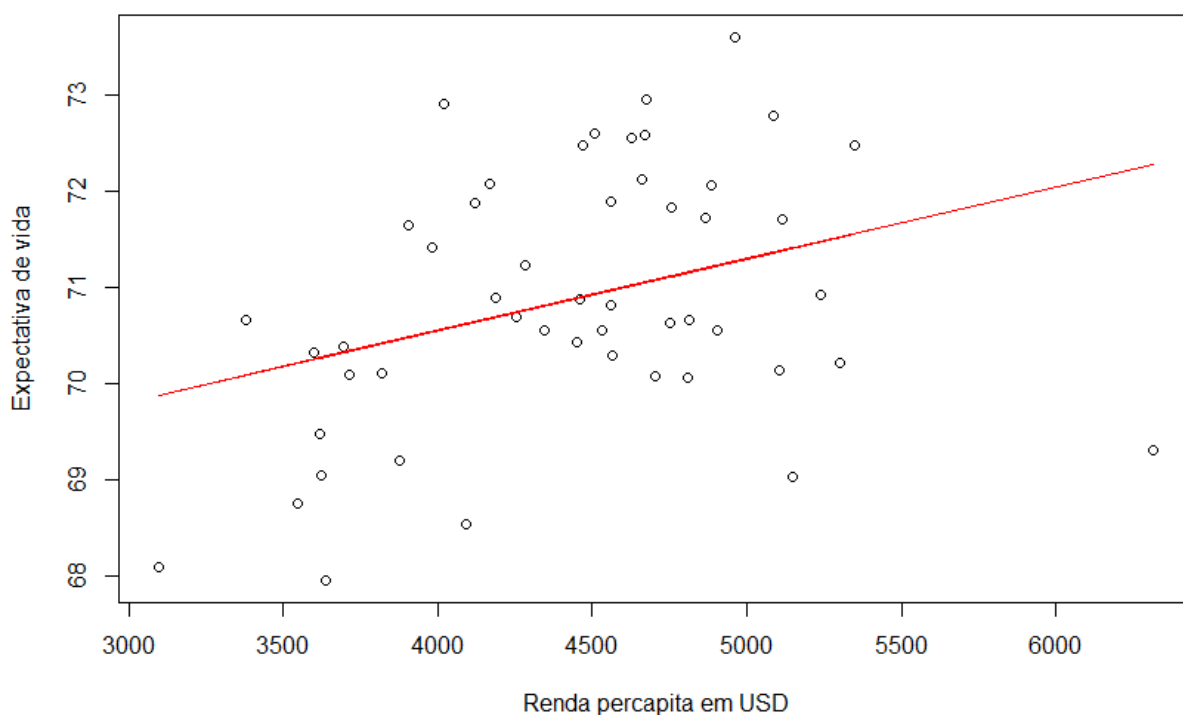


Gráfico 11.1. Gráfico de dispersão com a reta estimada entre a expectativa de vida e a renda per capita.

A partir do gráfico de dispersão (Gráfico 11.1), observamos que a expectativa de vida aumenta conforme o aumento da renda per capita da população. Assim, há evidências de uma relação linear crescente entre as variáveis, a qual é expressa pela seguinte equação estimada

$$\hat{y} = (6,758 \times 10) + (7,433 \times 10^{-4})x,$$

em que \hat{y} representa a expectativa de vida média dada uma renda per capita x da população.

Nesse íterim, observa-se dois parâmetros, β_0 que representa o valor esperado da expectativa de vida média quando a renda per capita é zero e β_1 é a mudança na média dessa expectativa de vida quando aumenta-se uma unidade no valor da renda per capita.

A adequabilidade da hipótese de que os erros do modelo linear simples são distribuídos normalmente com variância constante pode ser verificada nos Gráficos 12.1 e 13.1.

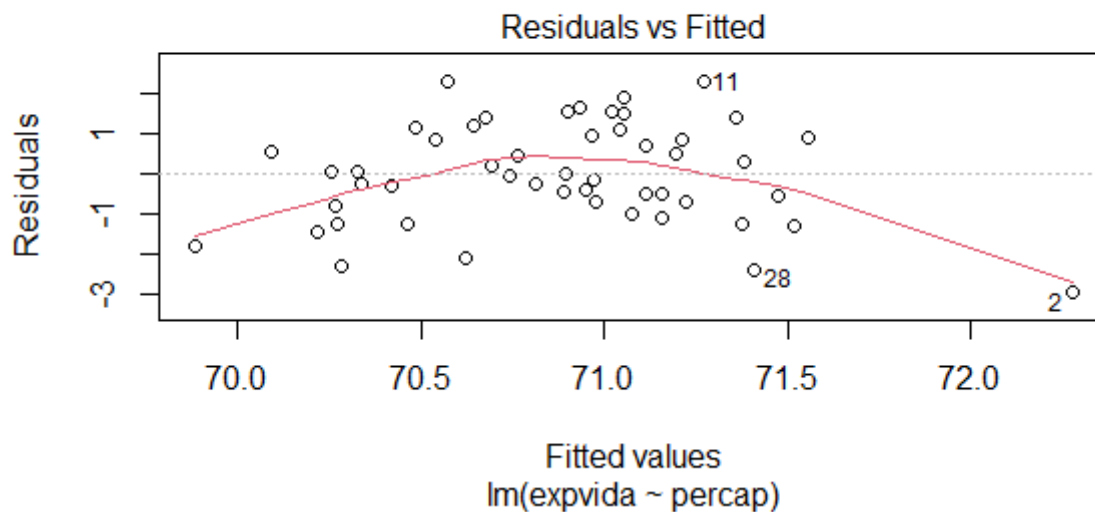


Gráfico 12.1. Gráfico de variância constante entre expectativa de vida e renda per capita

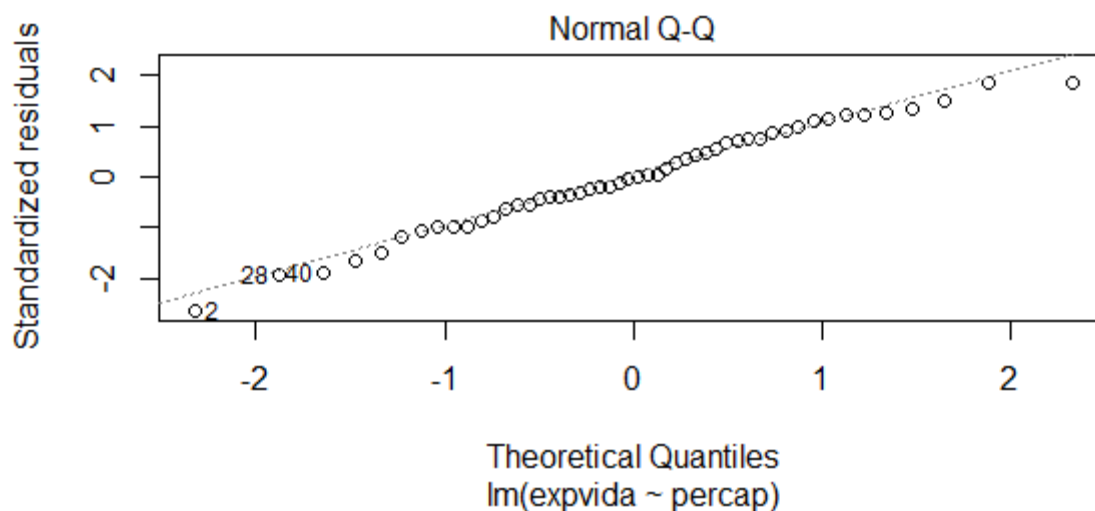


Gráfico 13.1. Gráfico de normalidade entre expectativa de vida e renda per capita

Levando-se em conta que na análise bidimensional, apesar de apresentar uma taxa de correlação positiva baixa (de 0,34) entre renda per capita e expectativa de vida, têm-se com a regressão linear, possuindo hipóteses confirmadas de que os erros do modelo linear simples são distribuídos normalmente com variância constante, uma reiteração de resultados.

Apesar do gráfico de variância constante entre expectativa de vida e renda per capita (Gráfico 12.1) não ter obtido um resultado regular, devido à ocorrência de outliers nas pontas,

isso não se torna um fator tão importante a ponto de rejeitar a hipótese de linearidade da correlação em questão.

2. Expectativa de vida e proporção de analfabetos:

Os resultados dos testes de hipótese para as variáveis β_0 e β_1 são apresentados na Figura 10.2.

```
Residuals:
    Min       1Q   Median       3Q      Max
-2.7169 -0.8063 -0.0349  0.7674  3.6675

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  72.3949     0.3383  213.973  < 2e-16 ***
analf        -1.2960     0.2570   -5.043  6.97e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figura 10.2. Resultado dos testes de hipótese de regressão linear entre expectativa de vida e proporção de analfabetos.

Analisando os testes de hipótese entre a expectativa de vida e a proporção de analfabetos (Figura 10.2), observamos que é razoável assumir que existe uma relação linear entre as variáveis em questão com um nível de significância de 0,1%.

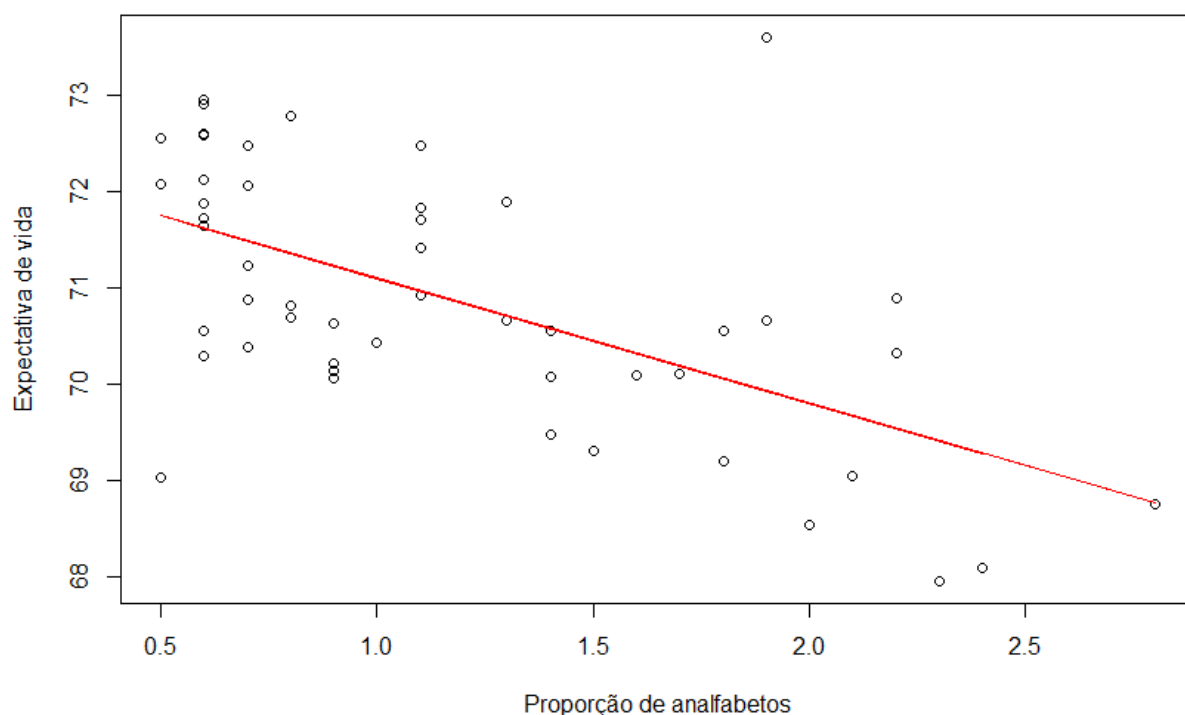


Gráfico 11.2. Gráfico de dispersão com a reta estimada entre expectativa de vida e proporção de analfabetos.

A partir do gráfico de dispersão (Gráfico 11.2), observamos que a expectativa de vida *diminui* conforme aumenta-se a proporção de analfabetos da população. Assim, há evidências de uma relação linear decrescente entre as variáveis, a qual é expressa pela seguinte equação estimada

$$\hat{y} = 72,395 - 1,296x,$$

em que \hat{y} representa a expectativa de vida média dada a proporção de analfabetos x da população.

Nesse íterim, observa-se dois parâmetros, β_0 que representa o valor esperado da expectativa de vida média quando a proporção de analfabetos é zero e β_1 é a mudança na média dessa expectativa de vida quando aumenta-se uma unidade no valor da proporção de analfabetos.

A adequabilidade da hipótese de que os erros do modelo linear simples são distribuídos normalmente com variância constante pode ser verificada nos Gráficos 12.2 e 13.2.

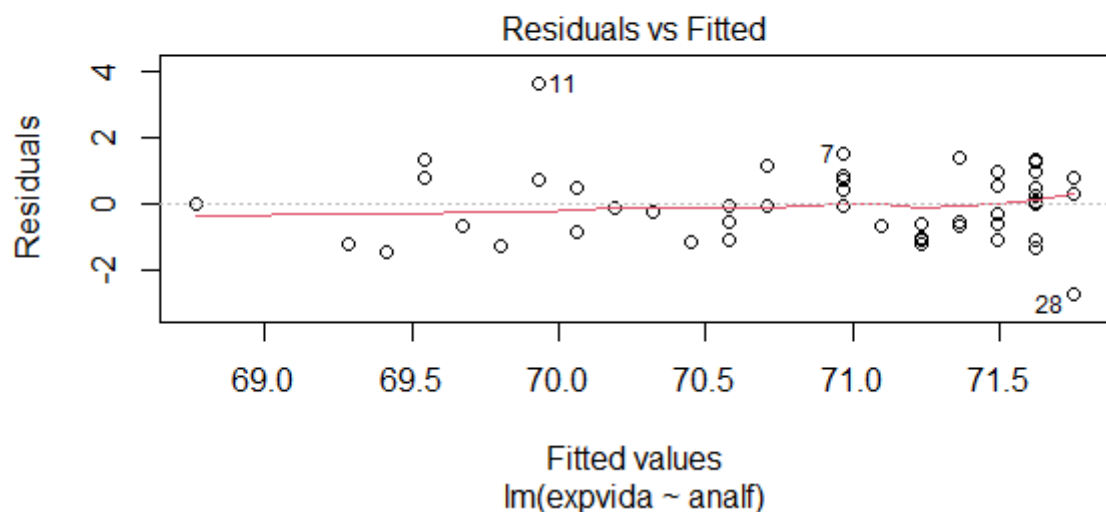


Gráfico 12.2. Gráfico de variância constante entre expectativa de vida e proporção de analfabetos.

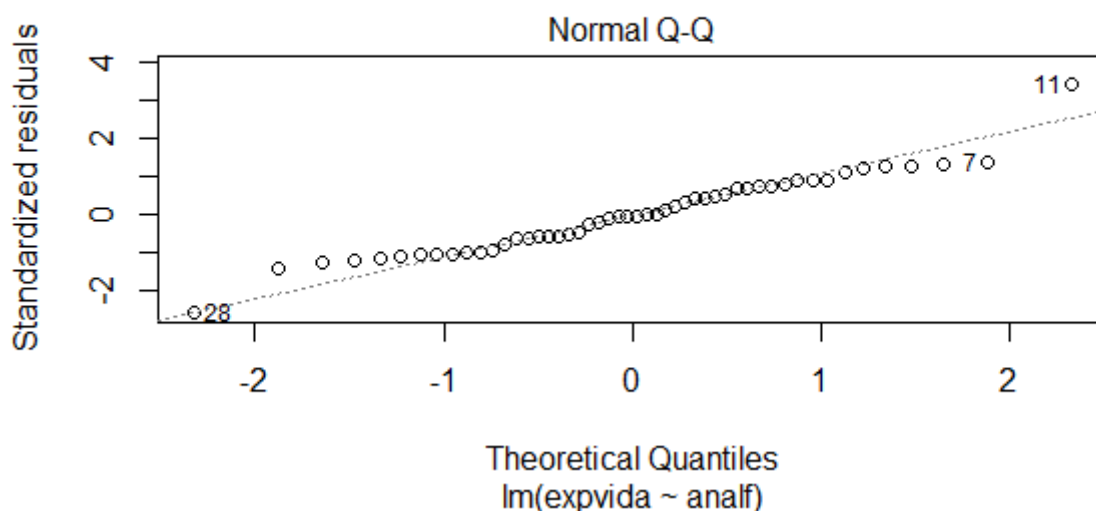


Gráfico 13.2. Gráfico de normalidade
entre expectativa de vida e proporção de analfabetos.

Levando-se em conta que a análise bidimensional, apresentou uma correlação negativa (-0,59) entre a proporção de analfabetos e expectativa de vida, têm-se com a regressão linear possuindo hipóteses confirmadas de que os erros do modelo linear simples são distribuídos normalmente com variância constante uma reiteração de resultados.

3. Expectativa de vida e taxa de criminalidade:

Os resultados dos testes de hipótese para as variáveis β_0 e β_1 são apresentados na Figura 10.3.

```
Residuals:
    Min       1Q   Median       3Q      Max
-1.81690 -0.48139  0.09591  0.39769  2.38691

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  72.97356    0.26997   270.30  < 2e-16 ***
crime       -0.28395    0.03279   -8.66  2.26e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figura 10.3. Resultado dos testes de hipótese de regressão linear entre expectativa de vida e taxa de criminalidade.

Analisando os testes de hipótese entre a expectativa de vida e a taxa de criminalidade (Figura 10.3), observamos que é razoável assumir que existe uma relação linear entre as variáveis em questão com um nível de significância de 0,1%.

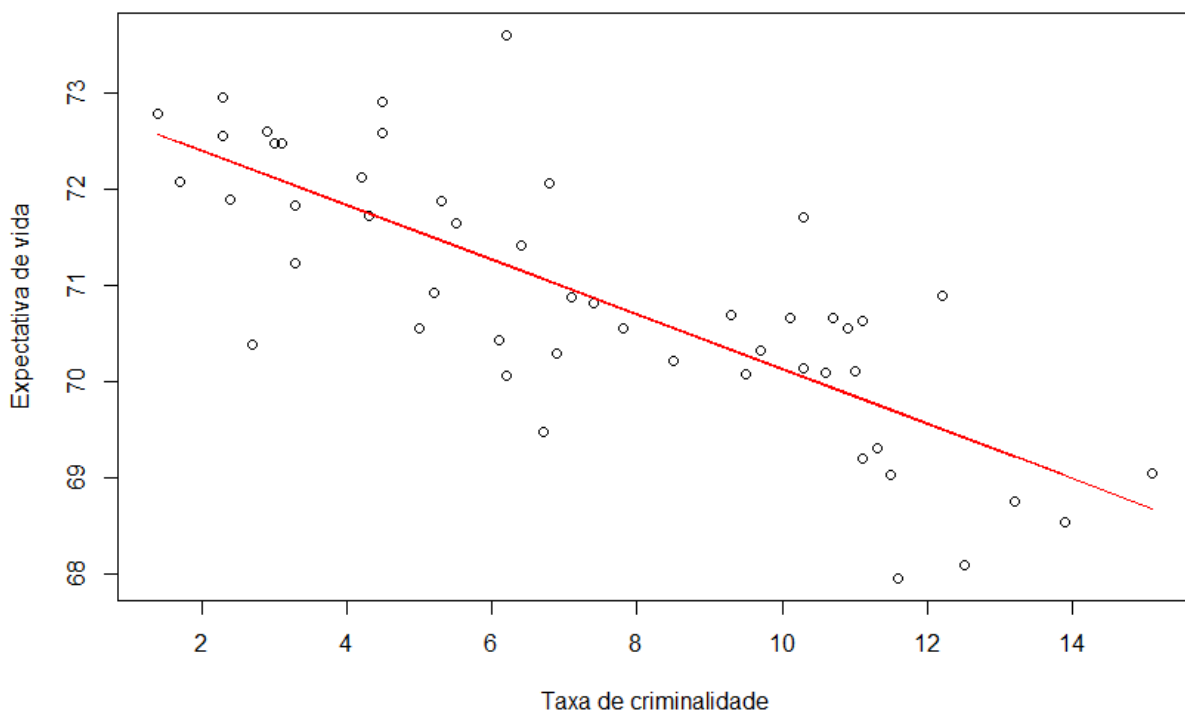


Gráfico 11.3. Gráfico de dispersão com a reta estimada entre expectativa de vida e taxa de criminalidade.

A partir do gráfico de dispersão (Gráfico 11.3), observamos que é razoável assumir que existe uma relação linear entre as variáveis em questão. Neste sentido, observamos que a expectativa de vida *diminui* conforme aumenta-se a taxa de criminalidade. Assim, há evidências de uma relação linear decrescente entre as variáveis, a qual é expressa pela seguinte equação estimada

$$\hat{y} = 72,973 - (2,839 \times 10^{-1}) x,$$

em que \hat{y} representa a expectativa de vida média dada a taxa de criminalidade x da população.

Nesse íterim, observa-se dois parâmetros, β_0 que representa o valor esperado da expectativa de vida média quando a taxa de criminalidade é zero e β_1 é a mudança na média dessa expectativa de vida quando aumenta-se uma unidade no valor da taxa de criminalidade.

A adequabilidade da hipótese de que os erros do modelo linear simples são distribuídos normalmente com variância constante pode ser verificada nos Gráficos 12.3 e 13.3.

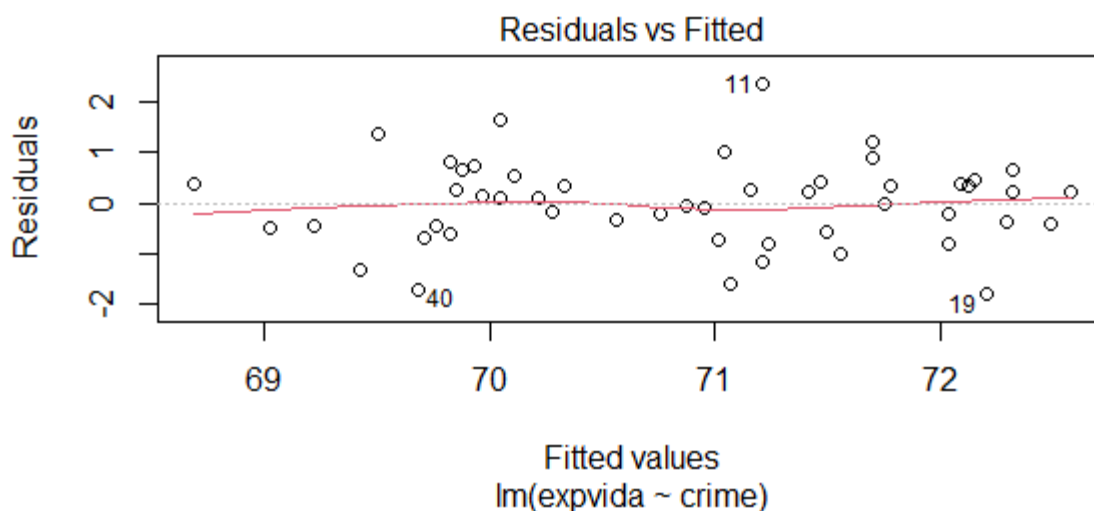


Gráfico 12.3. Gráfico de variância constante entre expectativa de vida e a taxa de criminalidade

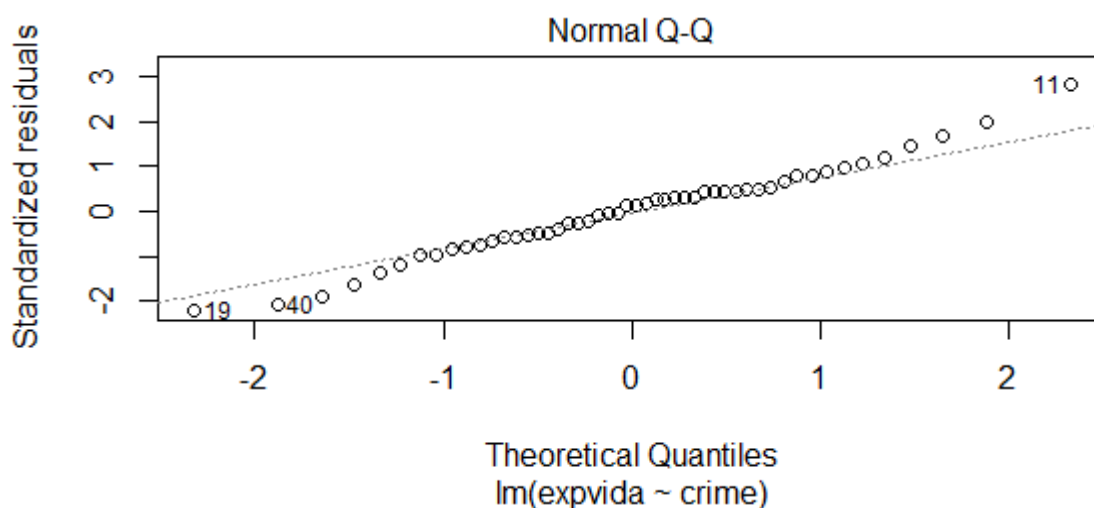


Gráfico 13.3. Gráfico de normalidade entre expectativa de vida e a taxa de criminalidade

Levando-se em conta que a análise bidimensional, apresentou uma correlação negativa forte (-0,78) entre taxa de criminalidade e expectativa de vida, têm-se com a regressão linear possuindo hipóteses confirmadas de que os erros do modelo linear simples são distribuídos normalmente com variância constante uma reiteração de resultados.

4. Expectativa de vida e porcentagem de estudantes com ensino de segundo grau completo:

Os resultados dos testes de hipótese para os coeficientes β_0 e β_1 são apresentados na Figura 10.4.

```

Residuals:
    Min       1Q   Median       3Q      Max
-3.01867 -0.67517 -0.07538  0.64483  2.17311

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  65.73965     1.04748   62.760  < 2e-16 ***
estud         0.09676     0.01950    4.961  9.2e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Figura 10.4. Resultado dos testes de hipótese de regressão linear entre expectativa de vida e a porcentagem de estudantes com segundo grau.

Analisando os testes de hipótese entre a expectativa de vida e a porcentagem de estudantes que completam o segundo grau (Figura 10.4), observamos que é razoável assumir que existe uma relação linear entre as variáveis em questão com um nível de significância de 0,1%.

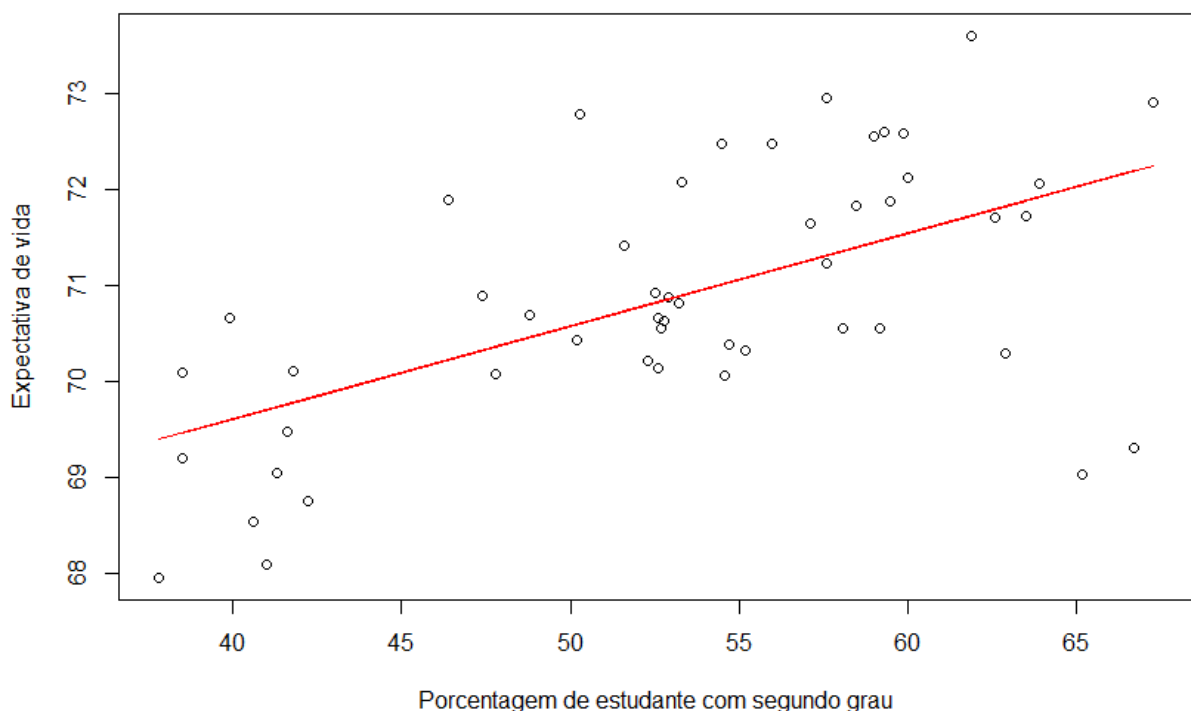


Gráfico 11.4. Gráfico de dispersão com a reta estimada entre expectativa de vida e porcentagem de estudantes com segundo grau.

A partir do gráfico de dispersão (Gráfico 11.4), observamos que a expectativa de vida aumenta conforme aumenta-se a porcentagem de estudantes com segundo grau completo. Assim, há evidências de uma relação linear crescente entre as variáveis, a qual é expressa pela seguinte equação estimada:

$$\hat{y} = 65,739 + (9,676 \times 10^{-2})x,$$

em que \hat{y} representa a expectativa de vida média dada a porcentagem de estudantes com o segundo grau completo x da população.

Nesse íterim, observa-se dois parâmetros, β_0 que representa o valor esperado da expectativa de vida média quando a porcentagem de estudantes com o segundo grau completo é zero e β_1 é a mudança na média dessa expectativa de vida quando aumenta-se uma unidade no valor da porcentagem de estudantes com o segundo grau completo.

A adequabilidade da hipótese de que os erros do modelo linear simples são distribuídos normalmente com variância constante pode ser verificada nos Gráficos 12.4 e 13.4.

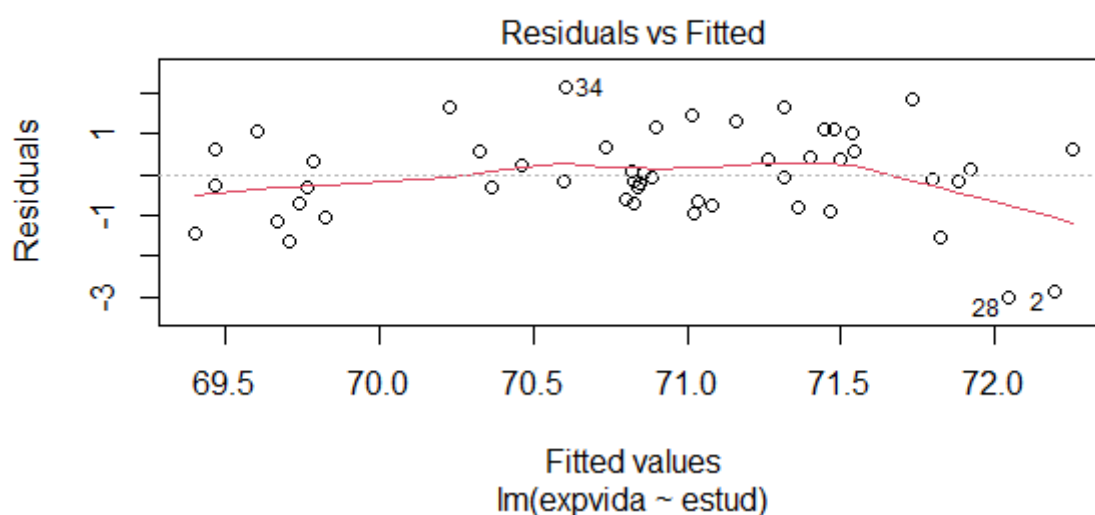


Gráfico 12.4 Gráfico de variância constante entre expectativa de vida e porcentagem de estudantes com ensino médio completo.

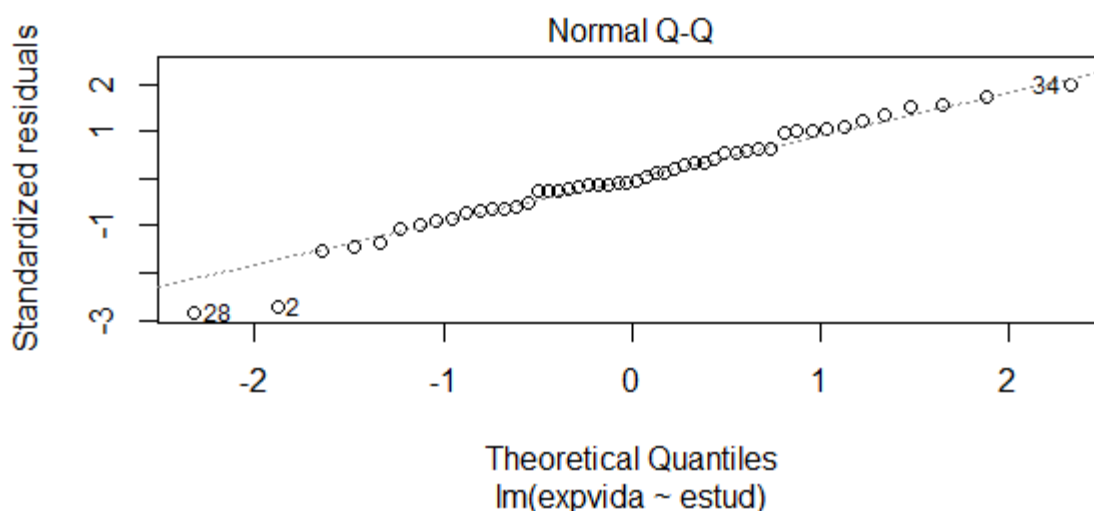


Figura 13.4. Gráfico de normalidade entre expectativa de vida e porcentagem de estudantes com ensino médio completo.

Levando-se em conta que a análise bidimensional, apresentou uma correlação positiva (0,58) entre porcentagem de estudantes com ensino médio completo e expectativa de vida, têm-se com a regressão linear possuindo hipóteses confirmadas de que os erros do modelo linear simples são distribuídos normalmente com variância constante uma reiteração de resultados.

5. Expectativa de vida e número de dias com temperaturas abaixo de zero grau Celsius:

Os resultados dos testes de hipótese para as variáveis β_0 e β_1 são apresentados na Figura 10.5.

```
Residuals:
    Min       1Q   Median       3Q      Max
-2.6515 -0.7852 -0.1183  0.9382  3.4284

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 70.171631   0.418883 167.521  <2e-16 ***
ndias        0.006768   0.003597   1.881   0.066 .
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figura 10.5. Resultado dos testes de hipótese de regressão linear entre expectativa de vida e o número de dias do ano com temperatura abaixo de zero.

Com base nos resultados dos testes de hipótese conduzidos, verifica-se que, apesar do teste para o coeficiente β_0 permitir a aceitação do modelo linear para um nível de significância de 0,1%, o teste para o coeficiente β_1 permite a aceitação do modelo linear apenas para um nível de significância de 10%. Com base nesses resultados, e levando em conta o nível de significância de 5% fixado para os testes conduzidos, conclui-se que não é razoável assumir uma relação linear entre as duas variáveis em questão.

Levando-se em conta que na análise bidimensional a taxa de correlação entre o número de dias do ano com temperatura abaixo de zero e a expectativa de vida indicou uma correlação baixa entre as duas variáveis (0,26), os testes de hipótese conduzidos reafirmam a baixa correlação entre as duas variáveis em questão.

6. Expectativa de vida e densidade demográfica:

Os resultados dos testes de hipótese para as variáveis β_0 e β_1 são apresentados na Figura 10.6.

```

Residuals:
    Min       1Q   Median       3Q      Max
-2.9063 -0.7295 -0.2602  1.0017  2.6765

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  71.0383     0.2039  348.407  <2e-16 ***
dens        -0.8498     0.4541   -1.871   0.0674 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Figura 10.6. Resultado dos testes de hipótese de regressão linear entre expectativa de vida e densidade populacional.

Analisando os testes de hipótese entre a expectativa de vida e a densidade populacional (Figura 10.6), observa-se que, assim como no caso do número de dias do ano com temperatura abaixo de zero, não é razoável assumir que existe uma relação linear entre as variáveis em questão. Apesar do coeficiente β_0 permitir a aceitação do modelo linear para um nível de significância de 0,1%, o teste para o coeficiente β_1 permite a aceitação do modelo linear apenas para um nível de significância de 10%, dessa forma, fixado o nível de significância de 5% para os testes, rejeita-se o modelo linear para as duas variáveis em questão.

Observa-se também que o coeficiente de correlação obtido na análise bidimensional das duas variáveis em questão ($-0,26$) é pouco significativo, fato este que reforça a não adoção do modelo linear e a baixa relação entre as duas variáveis.

5 CONSIDERAÇÕES FINAIS

Após a análise dos dados apresentados pela situação problema, no qual desejava-se explicar a expectativa de vida da população dos cinquenta (50) estados dos Estados Unidos através da renda per capita, taxa de analfabetismo, taxa de criminalidade, grau de estudos, dias do ano com temperaturas abaixo de 0°C e densidade populacional, obteve-se os seguintes resultados relevantes analisados abaixo.

Número de dias do ano com temperaturas abaixo de zero graus celsius e densidade populacional são fatores que possuem os menores impactos quanto à expectativa de vida populacional, como visto acima ambos casos não passaram nos testes de hipótese para verificar uma correlação linear entre cada variável e a expectativa de vida.

A baixa correlação entre renda per capita e a expectativa de vida da população se contrapõe com a ideia inicial do grupo que, através de uma análise simplista, possuía uma expectativa de correlação. Contudo, mesmo que ela tenha passado no teste de hipótese para ser considerada uma regressão linear, outros fatores indicam uma menor correlação entre as

variáveis: ela apresentou um coeficiente de correlação baixo de 0,34, apresentou um nível de significância maior que as outras que também passaram nos testes (de 5%) e teve o gráfico de variância constante fora de um resultado regular.

Além disso, a taxa de analfabetismo e a porcentagem de estudantes que concluem o ensino médio possuem semelhante correlação moderada. A primeira possui coeficiente negativo de -0,59, significando que quanto menor a população de analfabetos no local, maior a expectativa de vida da mesma. Concomitantemente, a segunda apresenta coeficiente positivo de 0,58, ou seja, quanto maior a população que possui grau de ensino médio completo, maior a expectativa de vida local.

Por fim, a maior correlação apresentada entre as características da população local e sua expectativa de vida foi a taxa de criminalidade do Estado, possuindo um coeficiente negativo de -0,78. Por conseguinte, quanto menor a taxa de criminalidade, o local em questão apresenta maior expectativa de vida.

Concomitantemente, ao realizar os testes de hipótese das correlações lineares, feitos através da regressão linear simples entre as variáveis, é possível observar uma reiteração de resultados entre a regressão e a análise bidimensional. Dessa forma, explícita-se que esta análise foi capaz de descrever uma relação entre as características observadas.

Dessa forma, conclui-se que a criminalidade é um fator determinante para analisar a expectativa de vida de uma população, possuindo uma relação inversamente proporcional que explica a problemática apresentada neste trabalho e, outrossim, a taxa de analfabetismo e a porcentagem de estudantes com ensino médio completo auxiliam, também, na interpretação desta como fatores secundários.

REFERÊNCIAS BIBLIOGRÁFICAS

BUSSAB, Wilton de O.; MORETTIN, Pedro A. **Estatística básica**. Saraiva, 2010