



Universidade de Brasília
Departamento de Estatística



Lista Prática 1

Letícia Inacio Rodrigues de Lima

Relatório apresentado para a disciplina
Análise de Regressão Linear - 2025.1 -
EST0038 como parte dos requisitos neces-
sários para aprovação.

Brasília
2025

Sumário

1	Introdução	3
2	Problema	4
2.1	Análise descritiva	4
2.1.1	Taxa de fecundidade	4
2.1.2	Expectativa de vida ao nascer	5
2.1.3	Taxa de analfabetismo adulto	5
2.1.4	Gasto privado com saúde em 2004	6
2.1.5	Gasto público com educação	7
2.1.6	Gasto com saúde per capita em 2004	8
2.1.7	Nascimentos assistidos por profissionais da saúde qualificados	9
2.1.8	Médicos por 100.000 habitantes	10
2.1.9	Prevalência de tabagismo (masculino)	11
2.1.10	Pesquisadores em Pesquisa e Desenvolvimento, por milhão de habitantes	12
2.1.11	Produto Interno Bruto, em milhões de dólares americanos	13
2.1.12	Legisladoras, altas funcionárias e gerentes	14
3	Transformações e ajuste de Modelos de Regressão Linear Simples	15
3.1	Taxa de fecundidade vs. Expectativa de vida	16
3.2	Taxa de Fecundidade vs. Taxa de analfabetismo adulto	16
3.3	Taxa de Fecundidade vs. Gasto com saúde per capita em 2004	18
3.4	Taxa de fecundidade vs. Nascimentos assistidos por profissionais de saúde qualificados	19
3.5	Taxa de fecundidade vs. Médicos por 100.000 habitantes	21
3.6	Taxa de fecundidade vs. Pesquisadores em Pesquisa e Desenvolvimento	22
4	Análises de diagnóstico	24
4.1	Normalidade dos erros	24
4.1.1	Taxa de fecundidade vs. Expectativa de vida	24
4.1.2	Taxa de fecundidade vs. Taxa de analfabetismo adulto	24
4.1.3	Taxa de fecundidade vs. Gasto com saúde per capita em 2004	25
4.1.4	Taxa de fecundidade vs. Nascimentos assistidos por profissionais de saúde qualificados	26
4.1.5	Taxa de fecundidade vs. Médicos por 100.000 habitantes	28
4.1.6	Taxa de fecundidade vs. pesquisadores em pesquisa e desenvolvimento	29
4.2	Homoscedasticidade dos erros	30
4.2.1	Taxa de fecundidade vs. Expectativa de vida	30
4.2.2	Taxa de fecundidade vs. Gasto com saúde per capita em 2004	31
4.2.3	Taxa de fecundidade vs. Nascimentos assistidos por profissionais de saúde qualificados	32
4.2.4	Taxa de fecundidade vs. Médicos por 100.000 habitantes	33
4.3	Não-correlação dos erros	34
4.3.1	Taxa de fecundidade vs. Nascimentos assistidos por profissionais de saúde qualificados	34
4.3.2	Taxa de fecundidade vs. Médicos por 100.000 habitantes	35
4.4	Alavancagem	36

4.4.1	Taxa de fecundidade vs. Nascimentos assistidos por profissionais de saúde qualificados	36
4.4.2	Taxa de fecundidade vs. Médicos por 100.000 habitantes	37
4.5	Influência	37
4.5.1	Taxa de fecundidade vs. Nascimentos assistidos por profissionais de saúde qualificados	38
4.5.2	Taxa de fecundidade vs. Médicos por 100.000 habitantes	39
5	Investigação de pontos atípicos	40
5.1	Taxa de fecundidade vs. Nascimentos assistidos por profissionais de saúde qualificados	41
5.2	Taxa de fecundidade vs. Médicos por 100.000 habitantes	42
6	Interpretação do modelo selecionado	42

1 Introdução

Este relatório tem o objetivo de solucionar um problema envolvendo análise de regressão linear simples. O problema envolve dados referentes a 185 países, contendo 15 variáveis, das quais serão utilizadas apenas 12. Sendo assim, a análise possui como objetivo o ajuste do melhor modelo de regressão linear simples possível para explicar a taxa de fecundidade, com base em uma das outras variáveis selecionadas, como expectativa de vida ao nascer (em anos), taxa de analfabetismo adulto (% da população com 15 anos ou mais), gasto privado com saúde em 2004 (% do PIB), gasto público com educação (% do PIB), gasto com saúde per capita em 2004 (PPA em dólares americanos), nascimentos assistidos por profissionais de saúde qualificados (%), médicos pro 100.000 habitantes, prevalência de tabagismo (masculino, % de adultos), pesquisadores em Pesquisa e Desenvolvimento (por milhão de habitantes), produto interno bruto (em bilhões de dólares americanos) e legisladoras, altas funcionárias e gerentes (% do sexo feminino).

2 Problema

2.1 Análise descritiva

Nesta seção, será realizada a análise descritiva das variáveis que foram previamente selecionadas com o objetivo de identificar o modelo de regressão linear simples que melhor explica a taxa de fecundidade.

2.1.1 Taxa de fecundidade

Neste tópico, será realizada a análise descritiva da variável taxa de fecundidade. Para tal, será utilizado um gráfico boxplot e uma tabela de medidas resumo.

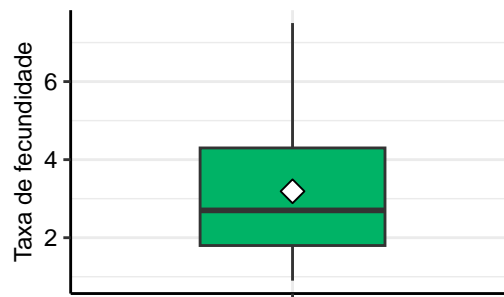


Figura 1: Boxplot

Tabela 1: Medidas resumo da Taxa de fecundidade

Estatística	Valor
Média	3.19
Desvio Padrão	1.71
Coef. de Variação	0.54
Mínimo	0.90
1º Quartil	1.80
Mediana	2.70
3º Quartil	4.30
Máximo	7.50

Analisando as informações expressas na Figura 1 e na Tabela 1, percebe-se que a média (3.19) assume um valor maior que a mediana (2.70), denotando a sensibilidade da média em relação aos valores extremos. Além disso, a amplitude da variável é de 6.6, uma vez que o maior valor é 7.5 e o menor é 0.9. É importante ressaltar que esta variável apresenta 4 dados faltantes, estes que foram retirados para que a análise fosse viabilizada.

2.1.2 Expectativa de vida ao nascer

Nesta seção, será realizada a análise descritiva da variável expectativa de vida ao nascer e a sua relação com a variável resposta taxa de fecundidade. Para tal, será utilizado um gráfico boxplot, um gráfico de dispersão com linha de regressão e uma tabela de medidas resumo.

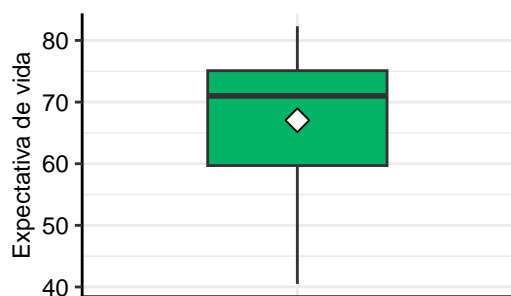


Figura 2: Boxplot Expectativa de vida

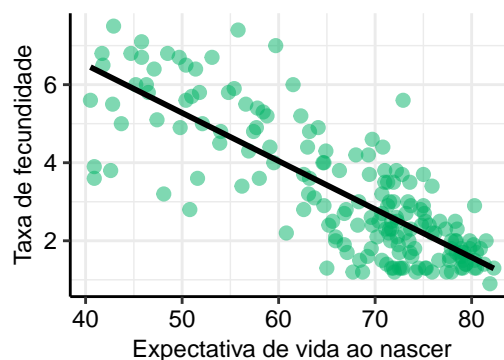


Figura 3: Relação entre Expectativa de vida e taxa de fecundidade

Tabela 2: Medidas resumo da Expectativa de vida ao nascer

Estatística	Valor
Média	67.05
Desvio Padrão	11.08
Coef. de Variação	0.17
Mínimo	40.50
1º Quartil	59.70
Mediana	71.00
3º Quartil	75.10
Máximo	82.30

Observando a Figura 2 e a Tabela 2, é visível que a média (67.05) é afetada por valores extremos, uma vez que se distancia levemente da mediana (71). Além disso, é possível perceber uma assimetria negativa, uma vez que 25% dos dados estão entre 71 e 75.10, podendo interferir nos resultados dos pressupostos da regressão linear simples, em especial o de normalidade dos erros. Ademais, analisando a relação entre as duas variáveis, expressa na Figura 3, percebe-se uma relação linear negativa, ou seja, quando menor a taxa de fecundidade, maior a expectativa de vida ao nascer.

2.1.3 Taxa de analfabetismo adulto

Neste tópico, será realizada a análise descritiva da variável taxa de analfabetismo adulto e a sua relação com a variável resposta taxa de fecundidade. Para tal, será utilizado um

gráfico boxplot, um gráfico de dispersão com linha de regressão e uma tabela de medidas resumo.

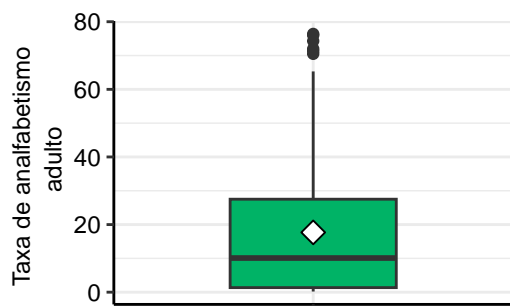


Figura 4: Boxplot Taxa de analfabetismo

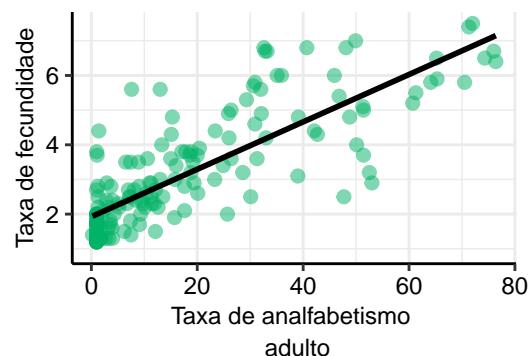


Figura 5: Relação entre Taxa de analfabetismo e Taxa de fecundidade

Tabela 3: Medidas resumo da Taxa de analfabetismo adulto

Estatística	Valor
Média	17.69
Desvio Padrão	19.86
Coef. de Variação	1.12
Mínimo	0.20
1º Quartil	1.35
Mediana	10.10
3º Quartil	27.50
Máximo	76.40

Observando a Figura 4 e a Tabela 3 é perceptível que a amplitude dos dados é alta, uma vez que o menor valor da amostra é 0.20 e o maior é 76.40. A média (17.69) é afetada por valores extremos, dado que é maior que a mediana (10.10). Também, é notória a presença de valores que se distanciam do padrão geral dos dados, ou seja, outliers. Além disso, também é possível perceber uma assimetria positiva, que pode afetar os resultados dos pressupostos necessários para a regressão linear, em especial, o de normalidade dos erros. Ademais, analisando relação entre as variáveis taxa de analfabetismo adulto e taxa de fecundidade através da Figura 5, pode-se dizer que existe uma relação linear positiva, ou seja, conforme a taxa de fecundidade cresce, maior a taxa de analfabetismo adulto. É importante ressaltar que esta variável apresenta 14 dados faltantes, estes que foram retirados para que a análise fosse viabilizada.

2.1.4 Gasto privado com saúde em 2004

Nesta seção, será realizada a análise descritiva da variável gasto privado com saúde em 2004 e a sua relação com a variável resposta taxa de fecundidade. Para tal, será utilizado

um gráfico boxplot, um gráfico de dispersão com linha de regressão e uma tabela de medidas resumo.

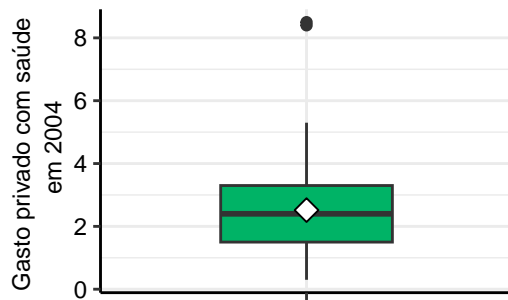


Figura 6: Boxplot Gasto privado com saúde em 2004

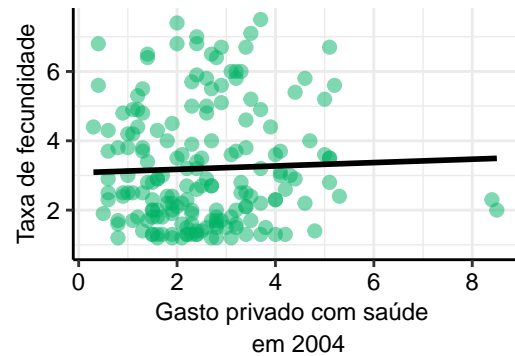


Figura 7: Relação entre Gasto privado com saúde em 2004 e taxa de fecundidade

Tabela 4: Medidas resumo do Gasto privado com saúde em 2004

Estatística	Valor
Média	2.52
Desvio Padrão	1.33
Coef. de Variação	0.53
Mínimo	0.30
1° Quartil	1.50
Mediana	2.40
3° Quartil	3.30
Máximo	8.50

Analisando a Figura 6 e a Tabela 4, percebe-se que a média (2.52) está próxima da mediana (2.40). Os dados aparentam ser simétricos, uma vez que a distância do 1° quartil para a mediana e da mediana para o 3° quartil é a mesma, de 0.9. Além disso, também é visível a presença de outliers. Ademais, observando a Figura 7, é possível dizer que não existe uma relação linear entre as variáveis gasto privado com saúde e taxa de fecundidade. É válido ressaltar que esta variável apresenta 1 dado faltante, este que foi retirado para que a análise fosse viabilizada.

2.1.5 Gasto público com educação

Nesta seção, será realizada a análise descritiva da variável gasto público com educação e a sua relação com a variável resposta taxa de fecundidade. Para tal, será utilizado um gráfico boxplot, um gráfico de dispersão com linha de regressão e uma tabela de medidas resumo.

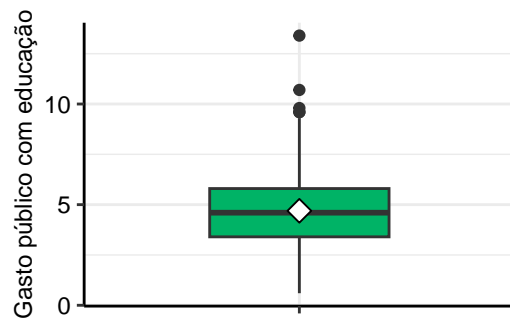


Figura 8: Boxplot Gasto público com educação

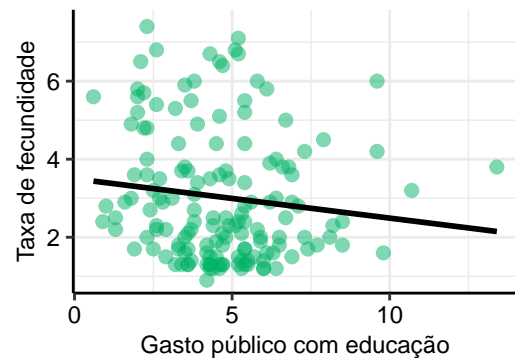


Figura 9: Relação entre Gasto público com educação e taxa de fecundidade

Tabela 5: Medidas resumo do Gasto público com educação

Estatística	Valor
Média	4.69
Desvio Padrão	2.05
Coef. de Variação	0.44
Mínimo	0.60
1º Quartil	3.40
Mediana	4.60
3º Quartil	5.80
Máximo	13.40

Observando a Figura 8 e a Tabela 5 pode-se perceber que a média (4.69) é próxima da mediana (4.60). Além disso, os dados aparentam ser simétricos, uma vez que a distância do 1º quartil para a mediana e da mediana para o 3º quartil é a mesma, de 1.20. Além disso, também é possível identificar a presença de pontos atípicos, conhecidos como outliers. Ademais, analisando a relação entre as variáveis gasto público com educação e taxa de fecundidade expressa pela Figura 9, conclui-se que não existe uma relação linear entre elas. É importante ressaltar que esta variável apresenta 28 dados faltantes, estes que foram retirados para que a análise fosse viabilizada.

2.1.6 Gasto com saúde per capita em 2004

Neste tópico, será realizada a análise descritiva da variável gasto com saúde per capita em 2004 e a sua relação com a variável resposta taxa de fecundidade. Para tal, será utilizado um gráfico boxplot, um gráfico de dispersão com linha de regressão e uma tabela de medidas resumo.

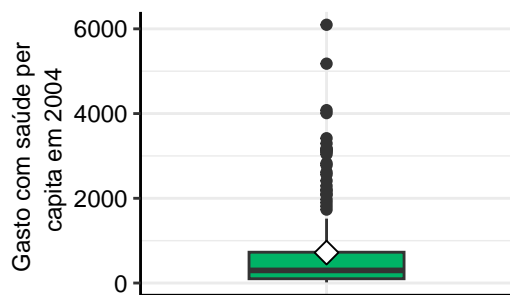


Figura 10: Boxplot Gasto com saúde per capita em 2004

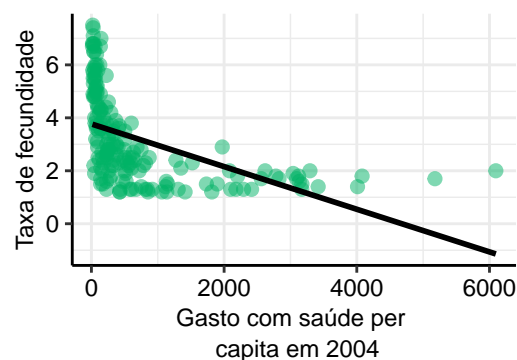


Figura 11: Relação entre Gasto com saúde per capita em 2004 e taxa de fecundidade

Tabela 6: Medidas resumo do Gasto com saúde per capita em 2004

Estatística	Valor
Média	718.01
Desvio Padrão	1037.01
Coef. de Variação	1.44
Mínimo	15.00
1º Quartil	100.50
Mediana	297.50
3º Quartil	727.00
Máximo	6096.00

Analisando a Figura 10 e a Tabela 6, é visível que a amplitude dos dados é alta (6081), uma vez que o menor valor é 15 e o maior é 6096. Além disso, a média (718.01) é influenciada pelos valores extremos, dado que se distancia, de forma considerável, da mediana (297.5). Também apresenta diversos outliers e uma assimetria positiva, fatores estes que podem influenciar os resultados dos pressupostos necessários para que um modelo de regressão linear seja válido. Ademais, observando a Figura 11, que apresenta a relação entre as variáveis gasto com saúde per capita em 2004 e taxa de fecundidade, é possível concluir que a relação não é linear, mas que, ao utilizar determinadas transformações, pode se aproximar da linearidade. É válido ressaltar que esta variável apresenta 5 dados faltantes, estes que foram retirados para que a análise fosse viabilizada.

2.1.7 Nascimentos assistidos por profissionais da saúde qualificados

Nesta seção, será realizada a análise descritiva da variável nascimentos assistidos por profissionais da saúde qualificados e a sua relação com a variável resposta. Para tal, será utilizado um gráfico boxplot, um gráfico de dispersão com linha de regressão e uma tabela de medidas resumo.

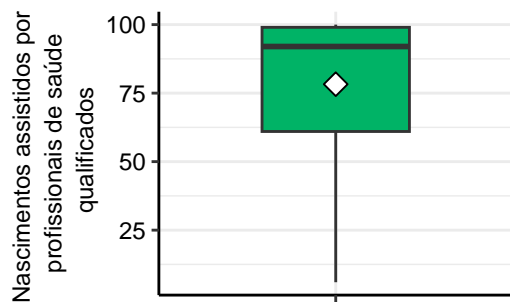


Figura 12: Boxplot Nascimentos assis-
tidos por profisionais da saúde qualifica-
dos

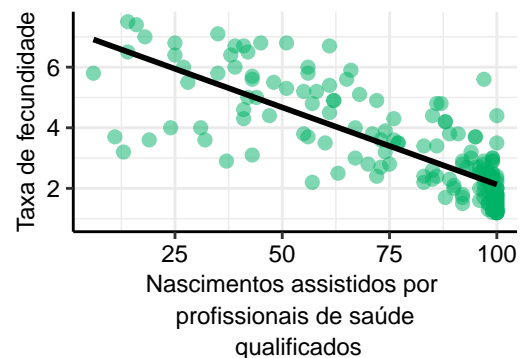


Figura 13: Relação entre Nascimentos
assisidos por profisionais da saúde qua-
lificados e taxa de fecundidade

Tabela 7: Medidas resumo do Nascimentos assisidos por profisionais da saúde qualifica-
dos

Estatística	Valor
Média	78.25
Desvio Padrão	26.42
Coef. de Variação	0.34
Mínimo	6.00
1º Quartil	61.00
Mediana	92.00
3º Quartil	99.00
Máximo	100.00

Observando a Figura 12 e a Tabela 7, percebe-se que a média (78.25) é afetada por valores extremos, uma vez que é menor que a mediana (92). Além disso, a amplitude (94) dos dados é alta, dado que o menor valor é 6 e o maior é 100. Também observa-se uma assimetria negativa, podendo impactar nos resultados dos pressupostos necessários para a regressão linear. Analisando a Figura 13, que ilustra a relação entre as variáveis nascimentos assistidos por profissionais de saúde qualificados e taxa de fecundidade, pode-se perceber uma relação linear negativa, ou seja, quando menor a taxa de fecundidade, maior o número de nascimentos assistidos por profissionais de saúde qualificados. É importante ressaltar que esta variável apresenta 7 dados faltantes, estes que foram retirados para que a análise fosse viabilizada.

2.1.8 Médicos por 100.000 habitantes

Neste tópico, será realizada a análise descritiva da variável médicos por 100.000 habitantes e a sua relação com a variável resposta taxa de fecundidade. Para tal, será utilizado um gráfico boxplot, um gráfico de dispersão com linha de regressão e uma tabela de medidas resumo.

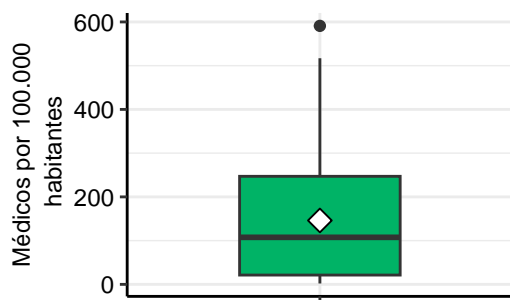


Figura 14: Boxplot Médicos por 100.000 habitantes

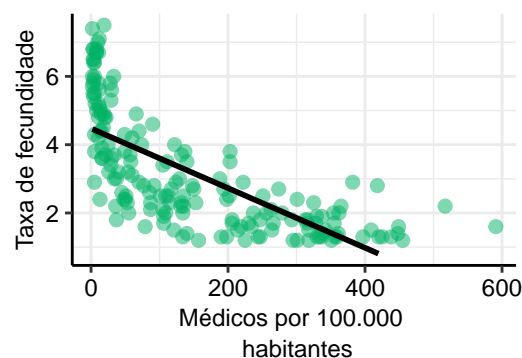


Figura 15: Relação entre Médicos por 100.000 habitantes e taxa de fecundidade

Tabela 8: Medidas resumo do Médicos por 100.000 habitantes

Estatística	Valor
Média	146.08
Desvio Padrão	138.55
Coef. de Variação	0.95
Mínimo	2.00
1º Quartil	21.25
Mediana	107.50
3º Quartil	247.00
Máximo	591.00

Analisando a Figura 14 e a Tabela 8, é visível que a média (146.08) é afetada por valores extremos, uma vez que é maior que a mediana (107.5). Além disso, observa-se a presença de um outlier e de uma assimetria positiva. Ambos estes fatores podem impactar os resultados dos pressupostos necessários para uma regressão linear. Destaca-se também que a amplitude (589) é alta, dado que o menor valor é 2 e o maior é 591. Ademais, observando a Figura 15, que ilustra a relação entre as variáveis médicos por 100.000 habitantes e taxa de fecundidade, não é possível enxergar uma linearidade. Entretanto, utilizar determinadas transformações podem aproximar essa relação da linearidade. É importante ressaltar que esta variável apresenta 3 dados faltantes, estes que foram retirados para que a análise fosse viabilizada.

2.1.9 Prevalência de tabagismo (masculino)

Neste seção, será realizada a análise descritiva da variável Prevalência de tabagismo (masculino) e a sua relação com a variável resposta taxa de fecundidade. Para tal, será utilizado um gráfico boxplot, um gráfico de dispersão com linha de regressão e uma tabela de medidas resumo.

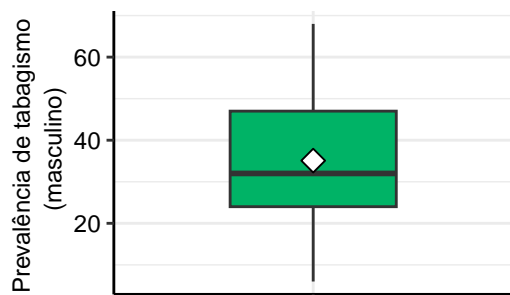


Figura 16: Boxplot Prevalência de tabagismo (masculino)

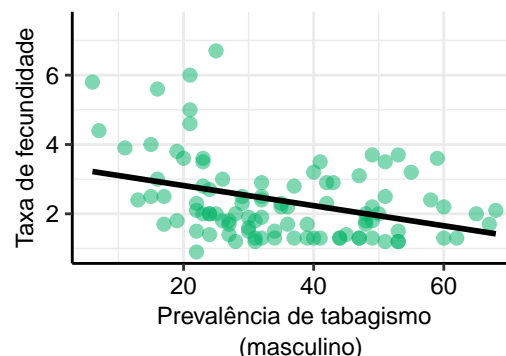


Figura 17: Relação entre Prevalência de tabagismo (masculino) e taxa de fecundidade

Tabela 9: Medidas resumo da Prevalência de tabagismo (masculino)

Estatística	Valor
Média	35.09
Desvio Padrão	14.40
Coef. de Variação	0.41
Mínimo	6.00
1º Quartil	24.00
Mediana	32.00
3º Quartil	47.00
Máximo	68.00

Analisando a Figura 16 e a Tabela 9, percebe-se que existe uma assimetria positiva, esta que pode impactar nos resultados dos pressupostos necessários para uma regressão linear. Ademais, a média (35.09) é maior que a mediana (32), denotando a sensibilidade aos valores extremos. Além disso, observando a Figura 17, que expressa a relação entre as variáveis prevalência de tabagismo e taxa de fecundidade, não é possível identificar uma relação linear entre as variáveis. É importante ressaltar que esta variável apresenta 88 dados faltantes, estes que foram retirados para que a análise fosse viabilizada.

2.1.10 Pesquisadores em Pesquisa e Desenvolvimento, por milhão de habitantes

Neste tópico, será realizada a análise descritiva da variável Pesquisadores em Pesquisa e Desenvolvimento, por milhão de habitantes e a sua relação com a variável resposta taxa de fecundidade. Para tal, será utilizado um gráfico boxplot, um gráfico de dispersão com linha de regressão e uma tabela de medidas resumo.

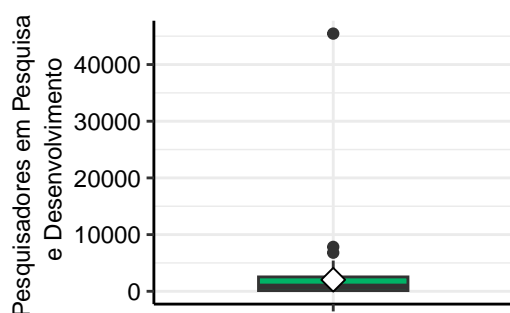


Figura 18: Boxplot Pesquisadores em Pesquisa e Desenvolvimento

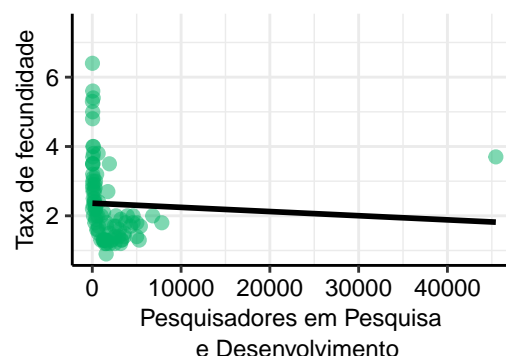


Figura 19: Relação entre Pesquisadores em Pesquisa e Desenvolvimento e

Tabela 10: Medidas resumo do Pesquisadores em Pesquisa e Desenvolvimento

Estatística	Valor
Média	2034.66
Desvio Padrão	4942.93
Coef. de Variação	2.43
Mínimo	15.00
1° Quartil	127.25
Mediana	848.00
3° Quartil	2538.00
Máximo	45 454.00

Observando a Figura 18 e a Tabela 10 é perceptível que existe um ponto que destoa do padrão dos dados de forma considerável, ponto este que pode estar influenciando o valor da média (2034.66), que é maior que a mediana (848). Percebe-se também uma assimetria positiva, uma vez que a distância entre o 1° quartil e a mediana é menor que a distância entre a mediana e o 3° quartil, sendo 720.75 e 1690, respectivamente. Além disso, analisando a Figura 19, que ilustra a relação entre as variáveis pesquisadores em Pesquisa e Desenvolvimento e taxa de fecundidade, não pode-se afirmar que existe linearidade, mas, utilizando as devidas transformações, esta relação pode se aproximar de uma relação linear. É importante ressaltar que esta variável apresenta 95 dados faltantes, estes que foram retirados para que a análise fosse viabilizada.

2.1.11 Produto Interno Bruto, em milhões de dólares americanos

Nesta seção, será realizada a análise descritiva da variável Produto Interno Bruto, em milhões de dólares americanos e a sua relação com a variável resposta taxa de fecundidade. Para tal, será utilizado um gráfico boxplot, um gráfico de dispersão com linha de regressão e uma tabela de medidas resumo.

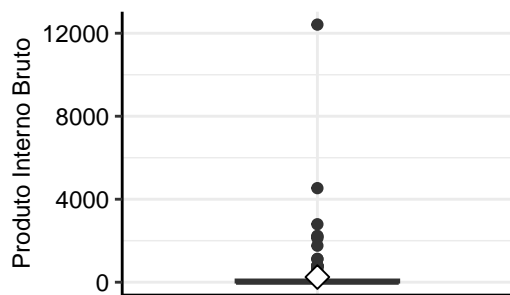


Figura 20: Boxplot Produto Interno Bruto

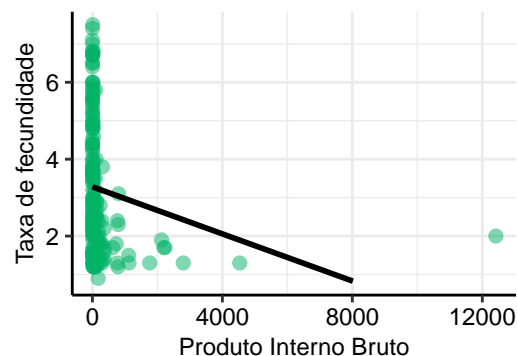


Figura 21: Relação entre Produto Interno Bruto e taxa de fecundidade

Tabela 11: Medidas resumo do Produto Interno Bruto

Estatística	Valor
Média	247.55
Desvio Padrão	1055.69
Coef. de Variação	4.26
Mínimo	0.10
1º Quartil	3.55
Mediana	14.20
3º Quartil	109.27
Máximo	12 416.50

Analisando a Figura 20 e a Tabela 11 é visível que os dados apresentam pontos discrepantes, estes que impactam o valor da média (247.55), que é maior que a mediana (14.2). Ademais, a amplitude dos dados (12416.4) é alta, dado que o menor valor é 0.1 e o maior é 12416.5. Percebe-se também uma assimetria positiva, uma vez que a distância entre o 1º quartil e a mediana é menor que a distância entre a mediana e o 3º quartil, sendo 10.65 e 95.07, respectivamente. Além disso, observando a relação entre as variáveis Produto Interno Bruto e taxa de fecundidade ilustrada na Figura 21, não é possível enxergar linearidade. É importante ressaltar que esta variável apresenta 7 dados faltantes, estes que foram retirados para que a análise fosse viabilizada.

2.1.12 Legisladoras, altas funcionárias e gerentes

Neste tópico, será realizada a análise descritiva da variável Legisladoras, altas funcionárias e gerentes e a sua relação com a variável resposta taxa de fecundidade. Para tal, será utilizado um gráfico boxplot, um gráfico de dispersão com linha de regressão e uma tabela de medidas resumo.

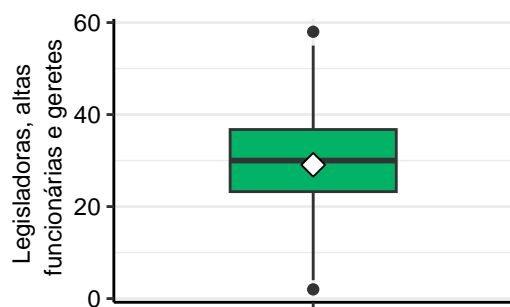


Figura 22: Boxplot Legisladoras, altas funcionárias e geretes

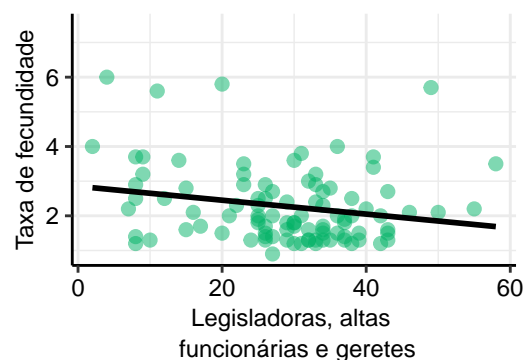


Figura 23: Relação entre Legisladoras, altas funcionárias e geretes e taxa de fecundidade

Tabela 12: Medidas resumo da Legisladoras, altas funcionárias e geretes

Estatística	Valor
Média	29.07
Desvio Padrão	11.71
Coef. de Variação	0.40
Mínimo	2.00
1º Quartil	23.25
Mediana	30.00
3º Quartil	36.75
Máximo	58.00

Observando a Figura 22 e a Tabela 12, é perceptível que os dados apresentam certa simetria, uma vez que a distância entre o 1º quartil e a mediana e entre a mediana e o 3º quartil é próxima, 6.25 e 6.75, respectivamente. Também é notória a presença de outliers, tanto superiores quanto inferiores. Além disso, analisando a Figura 23, que ilustra a relação entre as variáveis Legisladoras, altas funcionárias e geretes e taxa de fecundidade, não é possível identificar uma relação linear entre as duas variáveis. É válido ressaltar que esta variável apresenta 87 dados faltantes, estes que foram retirados para que a análise fosse viabilizada.

3 Transformações e ajuste de Modelos de Regressão Linear Simples

Foram realizados vários ajustes e transformações com as variáveis apresentadas anteriormente, resultando em doze modelos de regressão linear simples. As variáveis e transformações selecionadas para estes modelos foram escolhidas de acordo com a aproximação de uma relação linear com a variável resposta **taxa de fecundidade**.

3.1 Taxa de fecundidade vs. Expectativa de vida

Foi realizado o ajuste de um modelo de regressão linear simples para avaliar a relação entre a taxa de fecundidade e a expectativa de vida ao nascer.

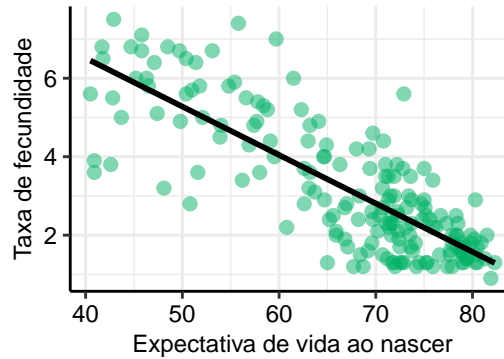


Figura 24: Modelo sem transformação

O modelo ilustrado pela Figura 24 não contém nenhuma transformação, pois a relação original entre as duas variáveis já apresenta certa linearidade. Os pontos apresentam uma relação negativa, mas estão consideravelmente dispersos ao redor da reta de regressão.

Tabela 13: Resumo do Modelo Ajustado Taxa de fecundidade vs. Expectativa de vida

Estatística	Valor
Intercepto (β_0)	11.448
Erro padrão do Intercepto	0.458
Coefficiente (β_1)	-0.123
Erro padrão do Coeficiente	0.007
R^2	0.651
R^2 ajustado	0.649
Erro padrão residual	1.012
p-valor do modelo	2.2×10^{-16}

A Tabela 13 apresenta as principais estatísticas do modelo ajustado. Dentre elas, destaca-se o coeficiente de determinação R^2 , que mede a proporção da variabilidade total da variável resposta explicada pelo modelo de regressão. O valor obtido foi $R^2 = 0.651$, ou seja, 65.1% da variação na variável resposta é explicada pelo modelo, o que é considerada uma explicação moderada.

3.2 Taxa de Fecundidade vs. Taxa de analfabetismo adulto

Foram realizados dois ajustes e transformações que resultaram em dois modelos de regressão linear simples para avaliar a relação entre a taxa de fecundidade e a taxa de analfabetismo adulto.

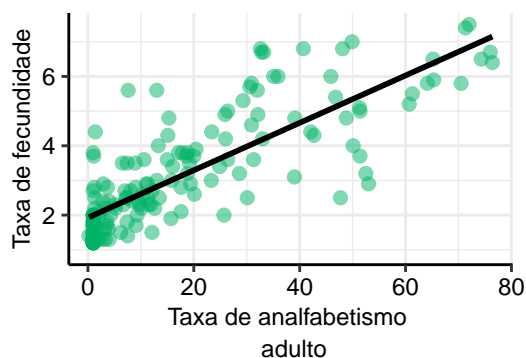


Figura 25: Modelo sem transformação

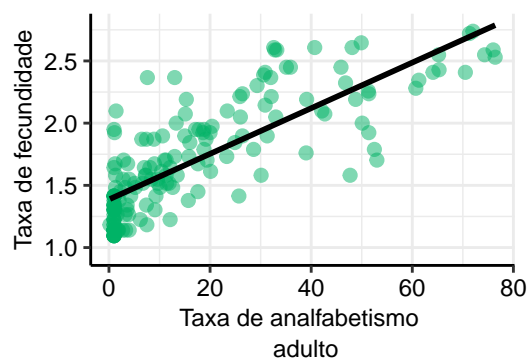


Figura 26: Modelo com transformação raiz quadrada

O modelo sem transformação, representado pela Figura 25, mostra que existe uma relação linear entre as duas variáveis. Também é perceptível que, apesar de existir uma relação positiva, os pontos estão dispersos ao redor da reta de regressão. Já o modelo representado na Figura 26 apresenta uma transformação raiz quadrática na variável taxa de fecundidade. Este modelo também apresenta uma relação linear positiva entre as duas variáveis, mas a dispersão dos pontos ao redor da reta de regressão é menor, fator que sugere um ajuste mais consistente em comparação com o modelo sem transformação.

Tabela 14: Resumo dos Modelos Ajustados para Taxa de fecundidade vs. Taxa de Analfabetismo Adulto

(a) Modelo sem transformação		(b) Modelo com transformação raiz quadrada	
Estatística	Valor	Estatística	Valor
Intercepto (β_0)	1.926	Intercepto (β_0)	1.387
Erro Padrão do Intercepto	0.102	Erro Padrão do Intercepto	0.029
Coeficiente (β_1)	0.068	Coeficiente (β_1)	0.018
Erro Padrão do Coeficiente	0.004	Erro Padrão do Coeficiente	0.001
R^2	0.662	R^2	0.643
R^2 Ajustado	0.660	R^2 Ajustado	0.641
Erro Padrão Residual	0.983	Erro Padrão Residual	0.274
p -valor do Modelo	2.2×10^{-16}	p -valor do Modelo	2.2×10^{-16}

A Tabela 14 a) apresenta as principais estatísticas do modelo sem transformação. O coeficiente de determinação referente a este ajuste é de $R^2 = 0.662$, denotando que 66.2% da variação na variável resposta é explicada pelo modelo. O erro padrão residual, medida que indica o quanto os valores previstos se desviam dos valores observados, é de 0.983. Já a Tabela 14 b) ilustra as informações referentes ao modelo ajustado com transformação de raiz quadrada. O coeficiente de determinação para esse modelo é de $R^2 = 0.643$, menor do que o modelo sem transformação. Entretanto, o erro padrão residual (0.274) para essa transformação é menor.

3.3 Taxa de Fecundidade vs. Gasto com saúde per capita em 2004

Foram realizados dois ajustes e transformações que resultaram em dois modelos de regressão linear simples para avaliar a relação entre a taxa de fecundidade e o gasto com saúde per capita em 2004, um com logaritmo na covariável gasto com saúde e outro com transformação logarítmica e com logaritmo na covariável gasto com saúde.

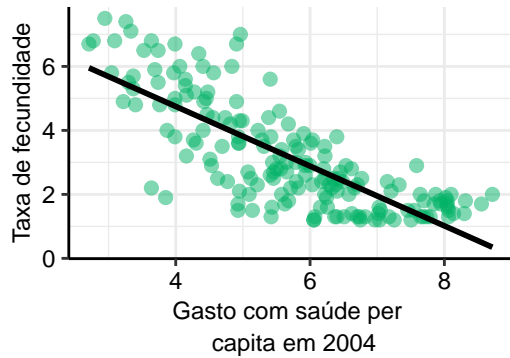


Figura 27: Modelo $\log(\text{Gasto com saúde})$

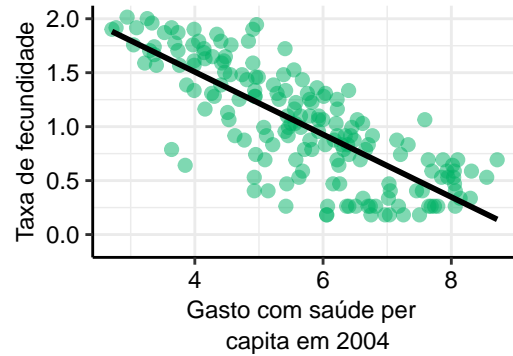


Figura 28: Modelo com transformação logarítmica e $\log(\text{Gasto com saúde})$

O modelo com logaritmo aplicado na variável gasto com saúde per capita, representado pela Figura 27 ilustra uma relação negativa que se aproxima da linearidade entre as variáveis. Além disso, a dispersão dos pontos ao redor da reta de regressão é considerável. A linearidade não descreve essa relação da melhor maneira. A Figura 28, por sua vez, ilustra o modelo com transformação logarítmica e com o logaritmo na variável gasto com saúde per capita. É visível que a dispersão dos pontos ao redor da reta de regressão é menor do que no modelo anterior. Ademais, a linearidade entre as duas variáveis ficou mais perceptível.

Tabela 15: Resumo dos Modelos Ajustados para Taxa de fecundidade vs. Gasto com Saúde per capita em 2004

(a) Modelo com $\log(\text{Gasto com saúde})$		(b) Modelo com transformação logarítmica e $\log(\text{Gasto com saúde})$	
Estatística	Valor	Estatística	Valor
Intercepto (β_0)	8.482	Intercepto (β_0)	2.667
Erro Padrão do Intercepto	0.329	Erro Padrão do Intercepto	0.103
Coefficiente (β_1)	-0.933	Coefficiente (β_1)	-0.290
Erro Padrão do Coeficiente	0.056	Erro Padrão do Coeficiente	0.018
R^2	0.612	R^2	0.608
R^2 Ajustado	0.610	R^2 Ajustado	0.606
Erro Padrão Residual	1.055	Erro Padrão Residual	0.331
p -valor do Modelo	2.2×10^{-16}	p -valor do Modelo	2.2×10^{-16}

A Tabela 15 a) mostra as principais estatísticas do modelo com logaritmo aplicado na variável gasto com saúde per capita. O coeficiente de determinação associado a este ajuste é de $R^2 = 0.612$, denotando que 61.2% da variação na variável resposta é explicada pelo modelo. O erro padrão residual é de 1.055, A Tabela 15 b), por sua vez, ilustra as informações referentes ao modelo ajustado com transformação logarítmica e logaritmo aplicado na variável gasto com saúde per capita. O coeficiente de determinação para esse modelo é de $R^2 = 0.608$, menor do que o modelo anterior. Mas, o erro padrão residual associado a essa transformação é de 0.331, que é menor quando comparado ao modelo sem transformação logarítmica na variável resposta.

3.4 Taxa de fecundidade vs. Nascimentos assistidos por profissionais de saúde qualificados

Foram realizados três ajustes e transformações que resultaram em três modelos de regressão linear simples para avaliar a relação entre a taxa de fecundidade e os nascimentos assistidos por profissionais de saúde qualificados, um sem transformação, outro com a transformação de Box-Cox e o último com transformação logarítmica.

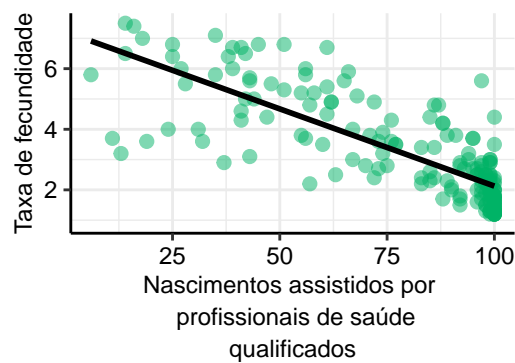


Figura 29: Modelo sem transformação

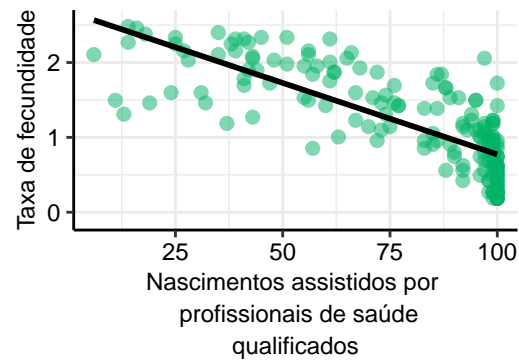


Figura 30: Modelo com transformação de Box-Cox

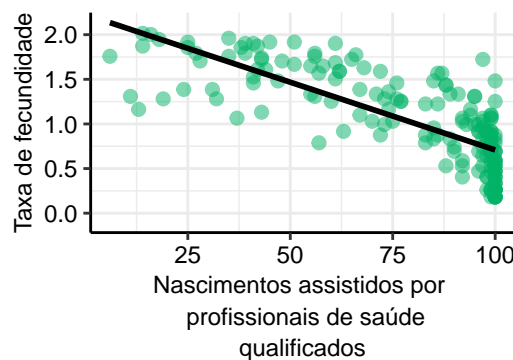


Figura 31: Modelo com transformação logarítmica

O modelo sem transformação, ilustrado pela Figura 29 apresenta uma relação linear negativa entre as duas variáveis. Também é visível que existe uma dispersão considerável dos pontos ao redor da reta da linha de regressão. A Figura 30 mostra o modelo com

transformação de Box-Cox. Com essa transformação, a relação também é negativa e a dispersão dos pontos parece diminuir. Também é perceptível que a linearidade entre as duas variáveis se intensifica. Por último, o modelo representado pela Figura 31 é o modelo com transformação logarítmica. Assim como nos outros dois modelos, a relação é negativa. A dispersão dos pontos em relação a reta de regressão diminui ainda mais e a linearidade entre as duas variáveis é ainda mais visível.

Tabela 16: Resumo dos Modelos Ajustados para Taxa de fecundidade vs. Nascimentos assistidos por profissionais de saúde qualificados

(a) Modelo sem transformação		(b) Modelo com transformação de Box-Cox	
Estatística	Valor	Estatística	Valor
Intercepto (β_0)	7.220	Intercepto (β_0)	2.682
Erro Padrão do Intercepto	0.245	Erro Padrão do Intercepto	0.097
Coefficiente (β_1)	-0.051	Coefficiente (β_1)	-0.019
Erro Padrão do Coeficiente	0.003	Erro Padrão do Coeficiente	0.001
R^2	0.629	R^2	0.603
R^2 Ajustado	0.627	R^2 Ajustado	0.600
Erro Padrão Residual	1.039	Erro Padrão Residual	0.412
p -valor do Modelo	2.2×10^{-16}	p -valor do Modelo	2.2×10^{-16}

(c) Modelo com transformação logarítmica	
Estatística	Valor
Intercepto (β_0)	2.226
Erro Padrão do Intercepto	0.079
Coefficiente (β_1)	-0.015
Erro Padrão do Coeficiente	0.001
R^2	0.589
R^2 Ajustado	0.586
Erro Padrão Residual	0.337
p -valor do Modelo	2.2×10^{-16}

A Tabela 16 a) ilustra as principais estatísticas do modelo sem transformação. O coeficiente de determinação associado a este ajuste é de $R^2 = 0.629$, denotando que 62.9% da variação na variável resposta é explicada pelo modelo. O erro padrão residual é de 1.039. A Tabela 16 b) ilustra as informações referentes ao modelo ajustado com transformação de Box-Cox. O coeficiente de determinação para esse modelo é de $R^2 =$

0.603, menor do que o modelo anterior. Entretanto, o erro padrão residual associado a essa transformação é de 0.412, que é menor quando comparado ao modelo sem transformação. Já o modelo com transformação logarítmica, representado pela Tabela 16 c), apresenta um coeficiente de determinação $R^2 = 0.589$, menor que os modelos apresentados anteriormente nesta seção. O erro padrão residual também é o menor entre os três modelos, de 0.337.

3.5 Taxa de fecundidade vs. Médicos por 100.000 habitantes

Foram realizados dois ajustes e transformações que resultaram em dois modelos de regressão linear simples para avaliar a relação entre a taxa de fecundidade e médicos por 100.000 habitantes, um com logaritmo na variável médicos por 100.000 habitantes e o outro com transformação logarítmica e com logaritmo na variável médicos por 100.000 habitantes.

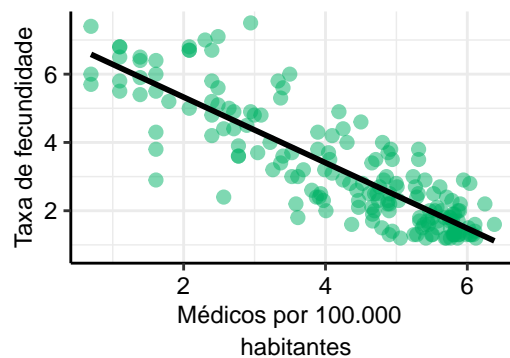


Figura 32: Modelo com $\log(\text{Médicos por } 100.00 \text{ hab.})$

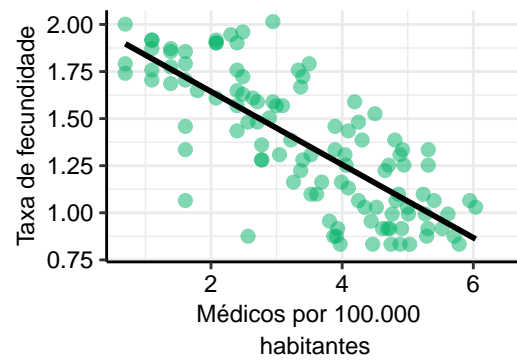


Figura 33: Modelo com transformação logarítmica e $\log(\text{Médicos por } 100.00 \text{ hab.})$

O modelo com logaritmo na variável médicos por 100.000 habitantes, representado na Figura 32 apresenta uma relação linear negativa entre as duas variáveis. Também é perceptível que existe uma certa dispersão dos pontos ao redor da reta da linha de regressão. Já a Figura 33 mostra o modelo com transformação logarítmica e com logaritmo na variável médicos por 100.000 habitantes. Essa transformação também apresenta uma relação negativa, mas a dispersão dos pontos parece aumentar. Comparando as duas transformações, a que apresenta a relação mais linear entre as duas variáveis é a representada pela Figura 33.

Tabela 17: Resumo dos Modelos Ajustados para Taxa de fecundidade vs. Médicos por 100.000 habitantes

(a) Modelo com $\log(\text{Médicos por } 100.00 \text{ hab.})$		(b) Modelo com transformação logarítmica e $\log(\text{Médicos por } 100.000 \text{ hab.})$	
Estatística	Valor	Estatística	Valor
Intercepto (β_0)	7.247	Intercepto (β_0)	2.259
Erro Padrão do Intercepto	0.196	Erro Padrão do Intercepto	0.064
Coeficiente (β_1)	-0.960	Coeficiente (β_1)	-0.293
Erro Padrão do Coeficiente	0.044	Erro Padrão do Coeficiente	0.014
R^2	0.733	R^2	0.707
R^2 Ajustado	0.732	R^2 Ajustado	0.705
Erro Padrão Residual	0.882	Erro Padrão Residual	0.287
p -valor do Modelo	2.2×10^{-16}	p -valor do Modelo	2.2×10^{-16}

A Tabela 17 a) mostra as principais estatísticas do modelo com logaritmo na variável médicos por 100.000 habitantes. O coeficiente de determinação deste ajuste é de $R^2 = 0.733$, denotando que 73.3% da variação na variável resposta é explicada pelo modelo. O erro padrão residual é de 0.882, A Tabela 17 b) ilustra as informações associadas ao modelo ajustado com transformação logarítmica e com logaritmo na variável médicos por 100.000 habitantes. O coeficiente de determinação deste modelo é de $R^2 = 0.707$, menor do que o modelo apresentado anteriormente. Mas, o erro padrão residual associado a essa transformação é de 0.287.

3.6 Taxa de fecundidade vs. Pesquisadores em Pesquisa e Desenvolvimento

Foram realizados dois ajustes e transformações que resultaram em dois modelos de regressão linear simples para avaliar a relação entre a taxa de fecundidade e pesquisadores em pesquisa e desenvolvimento, um com logaritmo na variável pesquisadores em pesquisa e desenvolvimento e o outro com transformação logarítmica e com logaritmo na variável pesquisadores em pesquisa e desenvolvimento.

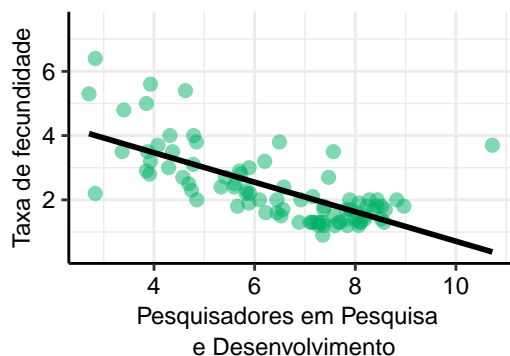


Figura 34: Modelo com $\log(\text{Pesq. em P\&D})$

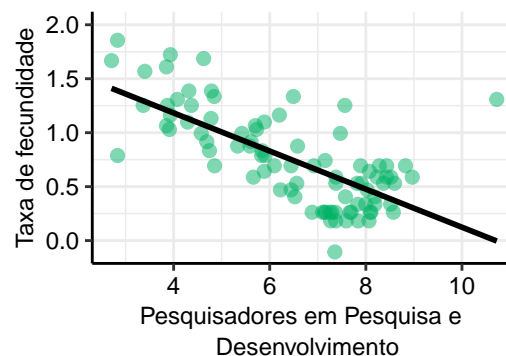


Figura 35: Modelo com transformação logarítmica e $\log(\text{Pesq. em P\&D})$

A Figura 34 representa o modelo com logaritmo na variável pesquisadores em pesquisa e desenvolvimento. Este apresenta uma relação negativa, e levemente linear entre as duas variáveis. Já a Figura 35 mostra o modelo com transformação logarítmica e com logaritmo na variável pesquisadores em pesquisa e desenvolvimento. Essa transformação também apresenta uma relação negativa, porém, apresenta mais linearidade entre as variáveis. Entretanto, a dispersão dos pontos ao redor da reta de regressão parece aumentar. Comparando as duas transformações, a que apresenta a relação mais linear entre as duas variáveis é a representada pela Figura 35.

Tabela 18: Resumo dos Modelos Ajustados para x Pesquisadores em Pesquisa e Desenvolvimento

(a) Modelo com $\log(\text{Pesq. em P\&D})$		(b) Modelo com transformação logarítmica e $\log(\text{Pesq. em P\&D})$	
Estatística	Valor	Estatística	Valor
Intercepto (β_0)	5.306	Intercepto (β_0)	1.889
Erro Padrão do Intercepto	0.345	Erro Padrão do Intercepto	0.130
Coeficiente (β_1)	-0.460	Coeficiente (β_1)	-0.176
Erro Padrão do Coeficiente	0.052	Erro Padrão do Coeficiente	0.019
R^2	0.477	R^2	0.487
R^2 Ajustado	0.471	R^2 Ajustado	0.481
Erro Padrão Residual	0.835	Erro Padrão Residual	0.315
p-valor do Modelo	6.864×10^{-14}	p-valor do Modelo	2.935×10^{-14}

A Tabela 18 a) ilustra as principais estatísticas do modelo com logaritmo na variável pesquisadores em pesquisa e desenvolvimento. O coeficiente de determinação associado a ajuste é de $R^2 = 0.477$, denotando que 47.7% da variação na variável resposta é explicada pelo modelo. O erro padrão residual é de 0.835. Já a Tabela 18 b) mostra as informações associadas ao modelo ajustado com transformação logarítmica e com logaritmo na variável pesquisadores em pesquisa e desenvolvimento. O coeficiente de determinação deste modelo

é de $R^2 = 0.481$, ou seja, 48.1% da variação na variável resposta é explicada pelo modelo, que é menor do que o modelo apresentado anteriormente. Entretanto, o erro padrão residual associado a essa transformação é de 0.315.

4 Análises de diagnóstico

Após ajustar os modelos de regressão, é necessário realizar uma análise de diagnóstico com a intenção de verificar se os pressupostos que garantem a veracidade estatística estão válidos. Entre estes pressupostos a serem analisados estão a normalidade, a homocedasticidade e a não-correlação dos erros. Além disso, também serão identificados possíveis pontos influentes ou discrepantes que podem impactar os ajustes. A seguir, serão apresentados gráficos e estatísticas utilizados para esta avaliação.

4.1 Normalidade dos erros

Neste tópico será avaliada a suposição de normalidade dos erros dos modelos, e para isso serão utilizados gráficos de envelope da distribuição normal e da distribuição t-student.

4.1.1 Taxa de fecundidade vs. Expectativa de vida

As figuras a seguir visam avaliar a normalidade dos erros e são referentes ao modelo ajustado para a taxa de fecundidade vs. expectativa de vida.

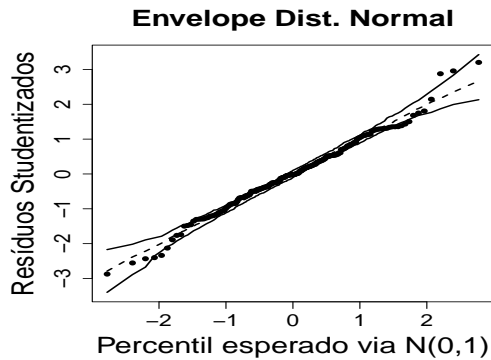


Figura 36: Modelo sem transformação

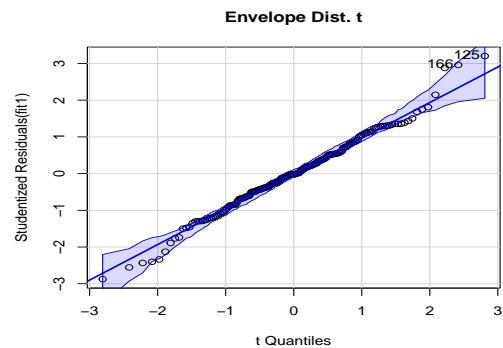


Figura 37: Modelo sem transformação

As Figuras 36 e 37 têm o objetivo de identificar a adequação do modelo ao pressuposto de normalidade dos erros. Como é possível observar, a maioria dos pontos estão dentro do envelope simulado, indicando que este modelo atende à este pressuposto.

4.1.2 Taxa de fecundidade vs. Taxa de analfabetismo adulto

As figuras a seguir visam avaliar a normalidade dos erros e são referentes aos modelos ajustados para a taxa de fecundidade vs. taxa de analfabetismo adulto e suas transformações.

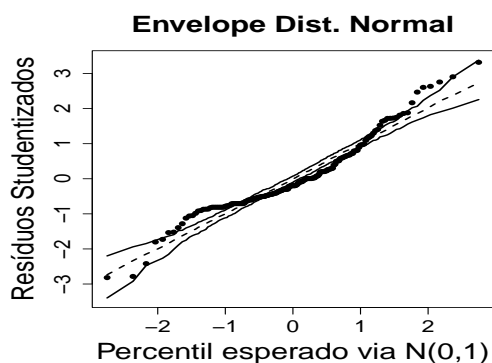


Figura 38: Modelo sem transformação

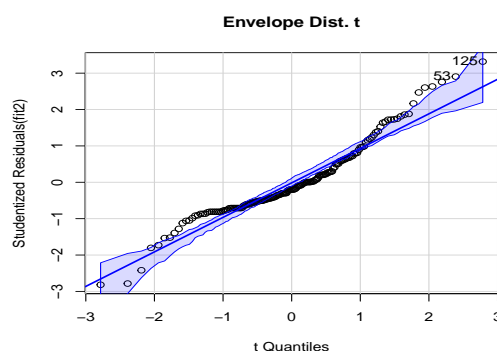


Figura 39: Modelo sem transformação

As Figuras 38 e 39 têm o objetivo de identificar a adequação do modelo sem transformação ao pressuposto de normalidade dos erros. É visível que existe uma quantidade razoável de pontos fora do envelope simulado. Este é um indicativo de desvio ao pressuposto.

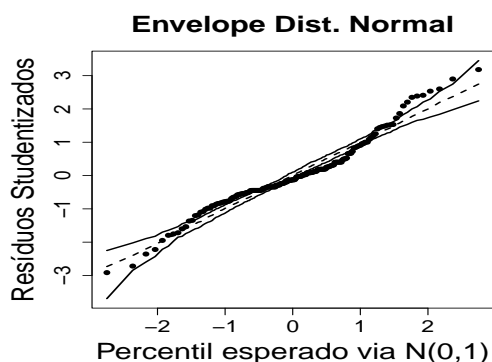


Figura 40: Modelo com transformação raiz quadratica

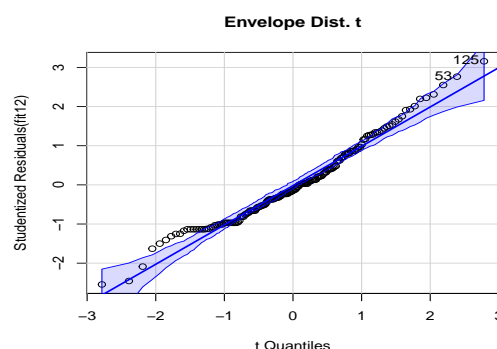


Figura 41: Modelo com transformação raiz quadratica

As Figuras 40 e 41 têm o objetivo de identificar a adequação do modelo com transformação raiz quadratica ao pressuposto de normalidade dos erros. Desse modo, é perceptível que a maioria das observações estão fora do envelope simulado. Assim como no modelo ajustado sem transformação, esse é indicativo de desvio do pressuposto de normalidade dos erros.

4.1.3 Taxa de fecundidade vs. Gasto com saúde per capita em 2004

As figuras a seguir têm o objetivo de avaliar a normalidade dos erros e são referentes aos modelos transformados ajustados para a taxa de fecundidade vs. gasto com saúde per capita em 2004.

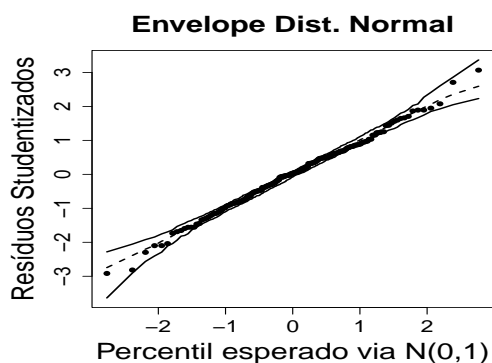


Figura 42: Modelo com $\log(\text{Gasto saúde})$

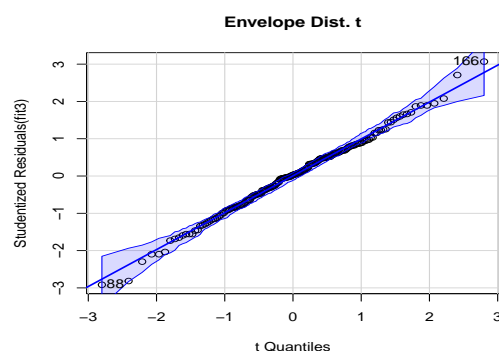


Figura 43: Modelo com $\log(\text{Gasto saúde})$

As Figuras 42 e 43 visam identificar a adequação do modelo com logaritmo no gasto com saúde per capita em 2004 ao pressuposto de normalidade dos erros. É visível que a maioria das observações estão localizadas dentro do envelope simulado, sendo este um indicativo de que não existem desvios da suposição de normalidade dos erros.

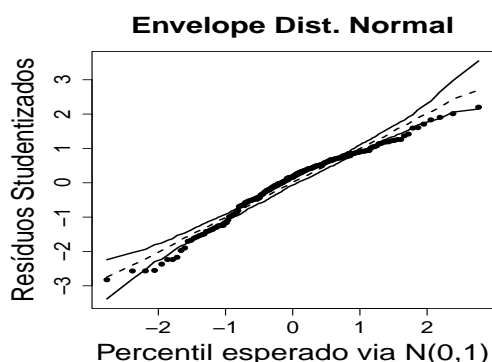


Figura 44: Modelo com transformação logarítmica e $\log(\text{Gasto com saúde})$

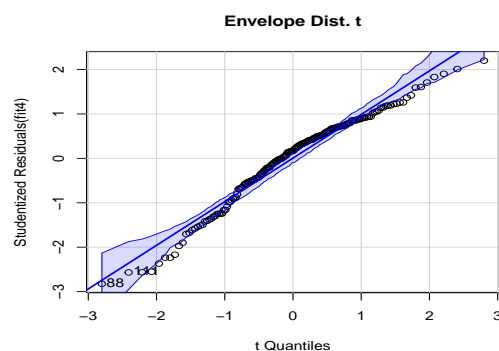


Figura 45: Modelo com transformação logarítmica e $\log(\text{Gasto com saúde})$

As Figuras 44 e 45 visam identificar a adequação do modelo com transformação logarítmica e com logaritmo na variável gasto com saúde per capita em 2004 ao pressuposto de normalidade dos erros. Dessa forma, é perceptível que a maioria dos pontos estão fora do envelope simulado, sendo este um indicativo de desvio ao pressuposto.

4.1.4 Taxa de fecundidade vs. Nascimentos assistidos por profissionais de saúde qualificados

As figuras a seguir têm o objetivo de avaliar a normalidade dos erros e são referentes aos modelos ajustados e suas transformações para a taxa de fecundidade vs. nascimentos assistidos por profissionais de saúde qualificados.

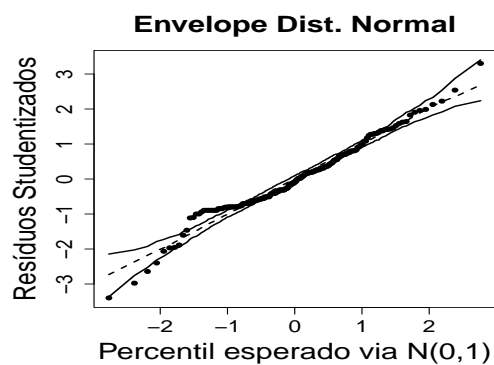


Figura 46: Modelo sem transformação

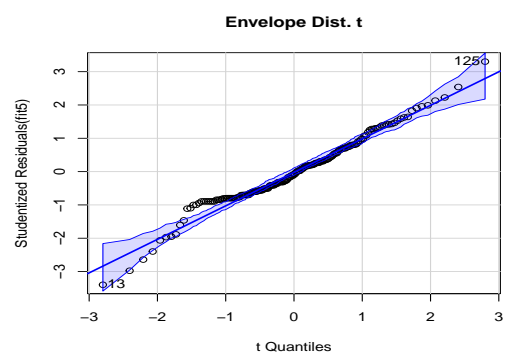


Figura 47: Modelo sem transformação

As Figuras 46 e 47 têm o objetivo de identificar a adequação do modelo sem transformação ao pressuposto de normalidade dos erros. Dessa forma, é perceptível que a maioria dos pontos estão fora do envelope simulado, sendo este um indicativo de desvio ao pressuposto.

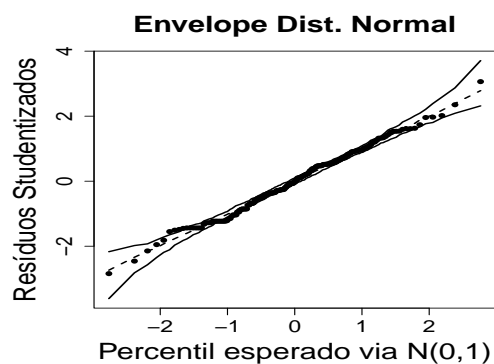


Figura 48: Modelo com transformação Box-Cox

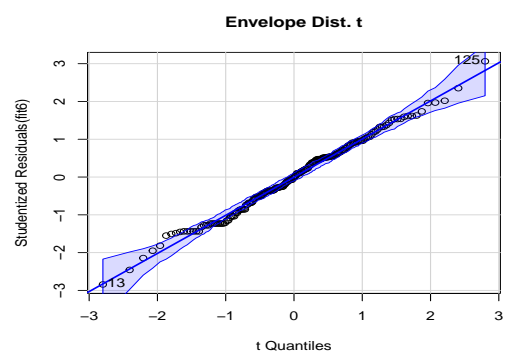


Figura 49: Modelo com transformação Box-Cox

As Figuras 48 e 49 tem o objetivo de identificar a adequação do modelo com transformação de Box-Cox ao pressuposto de normalidade dos erros. Sendo assim, pode-se perceber que a maioria das observações estão dentro do envelope simulado, indicando que não existe desvio aparente ao pressuposto de normalidade dos erros.

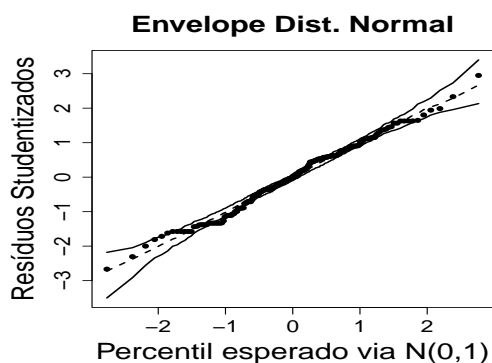


Figura 50: Modelo com transformação logarítmica

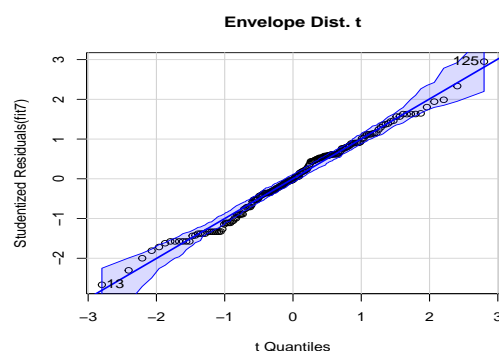


Figura 51: Modelo com transformação logarítmica

As Figuras 50 e 51 tem o objetivo de identificar a adequação do modelo com transformação logarítmica ao pressuposto de normalidade dos erros. Sendo assim, é visível que, apesar de alguns pontos estarem fora do envelope simulado, a maioria das observações estão dentro, o que é um indicativo de que não existe desvio aparente ao pressuposto de normalidade dos erros.

4.1.5 Taxa de fecundidade vs. Médicos por 100.000 habitantes

As figuras a seguir têm o objetivo de avaliar a normalidade dos erros e são referentes aos modelos transformados ajustados para a taxa de fecundidade vs. médicos por 100.000 habitantes.

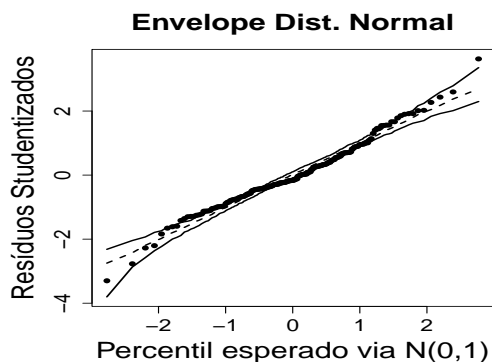


Figura 52: Modelo com $\log(\text{Médicos por } 100.000 \text{ hab.})$

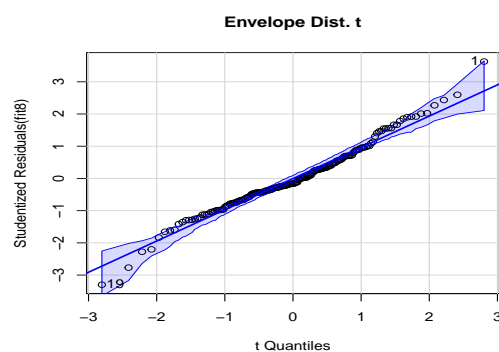


Figura 53: Modelo com $\log(\text{Médicos por } 100.000 \text{ hab.})$

As Figuras 52 e 53 visam identificar a adequação do modelo com logaritmo na variável médicos por 100.000 habitantes ao pressuposto de normalidade dos erros. Desse modo, é visível que a maioria das observações estão fora ou no limite do envelope simulado, indicando um desvio do pressuposto de normalidade dos erros.

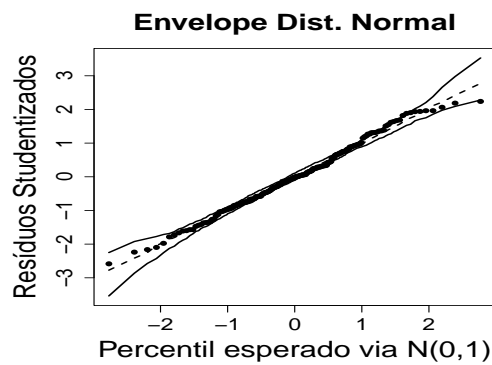


Figura 54: Modelo com transformação logarítmica e com $\log(\text{Médicos por } 100.000 \text{ hab.})$

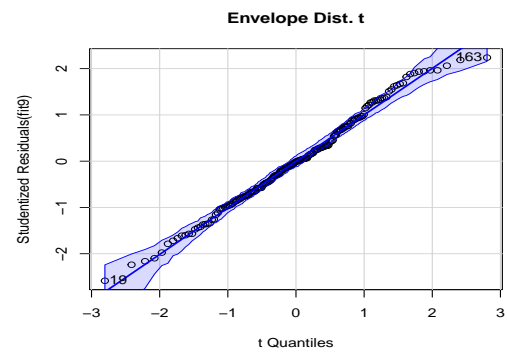


Figura 55: Modelo com transformação logarítmica e com $\log(\text{Médicos por } 100.000 \text{ hab.})$

As Figuras 54 e 55 tem o objetivo de identificar a adequação do modelo com transformação logarítmica e com logaritmo na variável médicos por 100.000 habitantes ao pressuposto de normalidade dos erros. Dessa forma, pode-se perceber que, apesar de alguns pontos estarem fora do envelope simulado, a maioria das observações estão dentro, o que é um indicativo de que não existe desvio aparente ao pressuposto de normalidade dos erros.

4.1.6 Taxa de fecundidade vs. pesquisadores em pesquisa e desenvolvimento

As figuras a seguir têm o objetivo de avaliar a normalidade dos erros e são referentes aos modelos transformados ajustados para a taxa de fecundidade vs. pesquisadores em pesquisa e desenvolvimento.

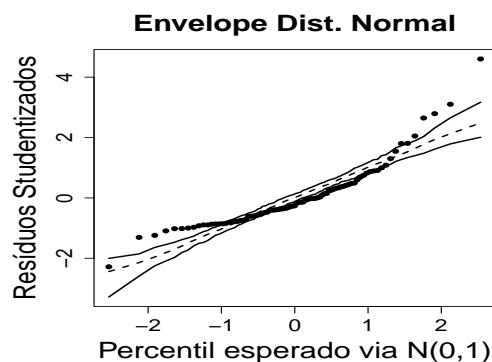


Figura 56: Modelo com $\log(\text{Pesq. em P\&D})$

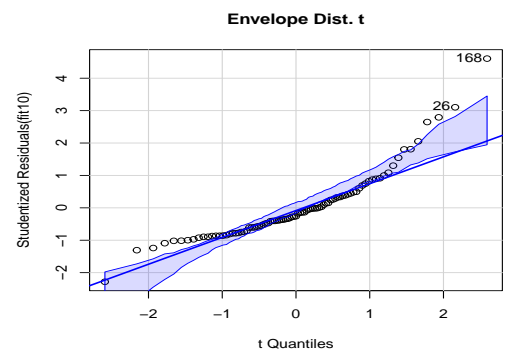


Figura 57: Modelo com $\log(\text{Pesq. em P\&D})$

As Figuras 56 e 57 tem o objetivo de identificar a adequação do modelo com logaritmo na variável pesquisadores em pesquisa e desenvolvimento ao pressuposto de normalidade dos erros. Sendo assim, pode-se perceber que a maioria das observações estão fora envelope simulado, fator que significa um desvio do pressuposto de normalidade dos erros.

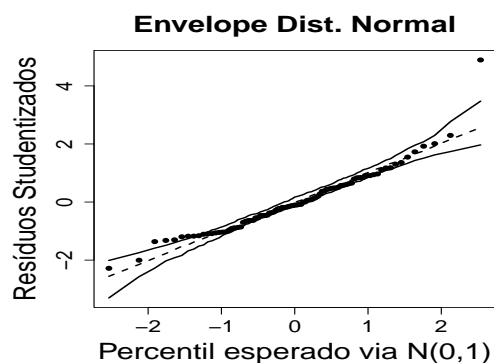


Figura 58: Modelo com transformação logarítmica e com $\log(\text{Pesq. em P\&D})$

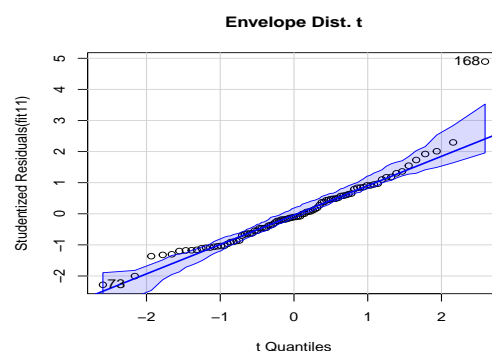


Figura 59: Modelo com transformação logarítmica e com $\log(\text{Pesq. em P\&D})$

As Figuras 58 e 59 visam identificar a adequação do modelo transformação logarítmica e com logaritmo na variável pesquisadores em pesquisa e desenvolvimento ao pressuposto de normalidade dos erros. Dessa forma, pode-se perceber que a maioria das observações estão fora ou no limite do envelope simulado, indicando um desvio do pressuposto de normalidade dos erros.

4.2 Homoscedasticidade dos erros

Neste tópico será realizada a avaliação da homoscedasticidade dos erros dos modelos, ou seja, da homogeneidade das variâncias dos erros, para aqueles modelos em que não houveram desvios no pressuposto de normalidade. Para isso, serão utilizados gráficos de resíduos studentizados vs. valores ajustados e o teste de Goldfeld-Quandt.

4.2.1 Taxa de fecundidade vs. Expectativa de vida

A figura e o teste a seguir visam avaliar a homoscedasticidade dos erros e fazem referência ao modelo ajustado para a taxa de fecundidade vs. expectativa de vida.

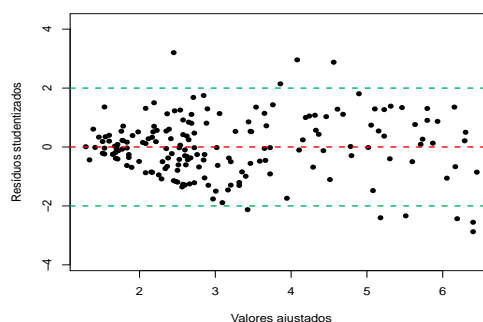


Figura 60: Modelo sem transformação

Observando a Figura 60, é possível identificar certo padrão na distribuição dos pontos ao longo do eixo dos valores ajustados. Sendo assim, existe uma suspeita de heterocedasticidade neste modelo. Para verificar esta suspeita, foi realizado um teste de hipóteses de Goldfeld-Quandt. As hipóteses a serem testadas são:

$$\begin{cases} H_0 : \text{ Os erros do modelo apresentam variância constante} \\ H_1 : \text{ Os erros do modelo não apresentam variância constante} \end{cases}$$

A interpretação do teste se baseia no p -valor, onde:

- $p > 0,05 \Rightarrow$ Não rejeitamos H_0 , ou seja, não há evidências de heterocedasticidade.
- $p \leq 0,05 \Rightarrow$ Rejeitamos H_0 , indicando heterocedasticidade significativa.

Tabela 19: Teste de homoscedasticidade dos erros do modelo

Modelo	p -valor
Sem transformação	3.53×10^{-8}

Como apresentado na Tabela 19, o p -valor do teste é < 0.001 , que é menor que o nível de significância estabelecido para o teste. Sendo assim, confirmando as suspeitas vistas na Figura 60, o modelo não é homoscedástico.

4.2.2 Taxa de fecundidade vs. Gasto com saúde per capita em 2004

A figura e o teste a seguir têm o objetivo de avaliar a homoscedasticidade dos erros e fazem referência ao modelo ajustado para a taxa de fecundidade vs. gasto com saúde per capita em 2004.

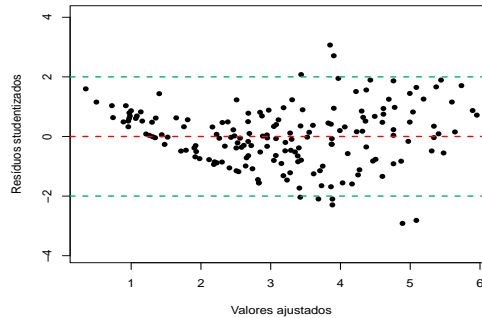


Figura 61: Modelo com $\log(\text{Gasto com saúde})$

Analisando a Figura 61, pode-se perceber um padrão de funil na distribuição dos pontos, ou seja, a variabilidade dos erros parece aumentar ao longo do eixo dos valores ajustados. Dessa forma, existe uma suspeita de heterocedasticidade neste modelo. Para verificar esta suspeita, foi realizado um teste de hipóteses de Goldfeld-Quandt. As hipóteses a serem testadas são:

$$\begin{cases} H_0 : \text{ Os erros do modelo apresentam variância constante} \\ H_1 : \text{ Os erros do modelo não apresentam variância constante} \end{cases}$$

A interpretação do teste se baseia no p -valor, onde:

- $p > 0,05 \Rightarrow$ Não rejeitamos H_0 , ou seja, não há evidências de heterocedasticidade.
- $p \leq 0,05 \Rightarrow$ Rejeitamos H_0 , indicando heterocedasticidade significativa.

Tabela 20: Teste de homoscedasticidade dos erros do modelo

Modelo	p -valor
$\log(\text{Gasto com saúde})$	7.955×10^{-10}

A Tabela 20 ilustra o p -valor associado ao teste realizado, este que é < 0.001 . Como o nível de significância estabelecido para o teste é de 5%, e portanto maior que o p -valor do teste, existem evidências para que a hipótese nula seja rejeitada. Desse modo, confirmando as suspeitas vistas na Figura 61, o modelo não é homoscedástico.

4.2.3 Taxa de fecundidade vs. Nascimentos assistidos por profissionais de saúde qualificados

As figuras e os testes a seguir têm o objetivo de avaliar a homoscedasticidade dos erros e fazem referência aos modelos transformados ajustados para a taxa de fecundidade vs. nascimentos assistidos por profissionais de saúde qualificados.

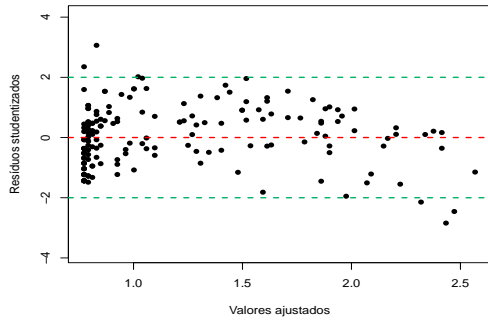


Figura 62: Modelo com transformação Box-Cox

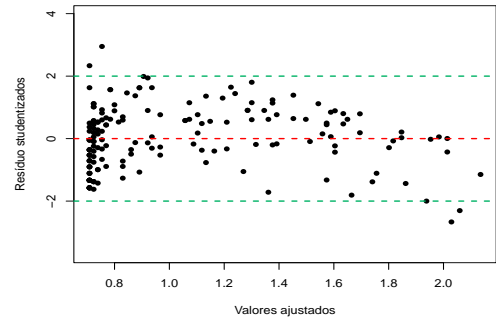


Figura 63: Modelo com transformação logarítmica

Observando as Figuras 62 e 63, não é possível identificar nenhum padrão na distribuição dos pontos ao longo do eixo dos valores ajustados, apesar de existir certa dispersão. Sendo assim, existe uma suspeita de homoscedasticidade em ambos os modelos ajustados. Para verificar estas suspeitas, foi realizado um teste de hipóteses de Goldfeld-Quandt. As hipóteses a serem testadas são:

$$\begin{cases} H_0 : \text{Os erros do modelo apresentam variância constante} \\ H_1 : \text{Os erros do modelo não apresentam variância constante} \end{cases}$$

A interpretação do teste se baseia no p -valor, onde:

- $p > 0,05 \Rightarrow$ Não rejeitamos H_0 , ou seja, não há evidências de heterocedasticidade.

- $p \leq 0,05 \Rightarrow$ Rejeitamos H_0 , indicando heterocedasticidade significativa.

Tabela 21: Testes de homoscedasticidade dos erros dos modelos

Modelo	p -valor
Transformação Box-Cox	0.261
Transformação logarítmica	0.975

A Tabela 20 apresenta os p -valores dos testes de hipóteses associados aos modelos com transformação de Box-Cox e com transformação logarítmica. Como o nível de significância estabelecido para os testes foi de 5%, e os p -valores apresentaram valores superiores, não existem evidências para rejeitar as hipóteses nulas de nenhum dos testes. Sendo assim, confirmando as suspeitas observadas na Figuras 62 e 63, os modelos são homoscedásticos.

4.2.4 Taxa de fecundidade vs. Médicos por 100.000 habitantes

A figura e o teste a seguir têm o objetivo de avaliar a homoscedasticidade dos erros e fazem referência ao modelo transformado ajustado para a taxa de fecundidade vs. médicos por 100.000 habitantes.

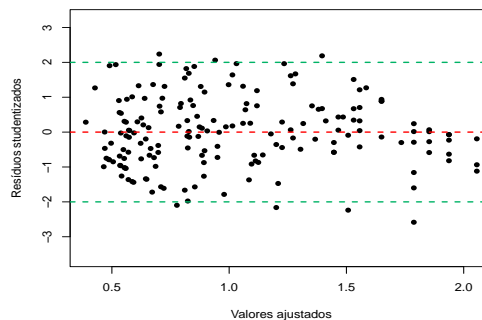


Figura 64: Modelo com transformação logarítmica e com $\log(\text{Médicos por } 100.000 \text{ hab.})$

Analisando a Figura 64, não é visível nenhum padrão na distribuição dos pontos ao longo do eixo dos valores ajustados. Além disso, a maioria dos pontos, que representam os erros, estão em torno do zero, que é o que se espera. Sendo assim, existe uma suspeita de homoscedasticidade neste modelo. Para verificar esta suspeita, foi realizado um teste de hipóteses de Goldfeld-Quandt. As hipóteses a serem testadas são:

$$\begin{cases} H_0 : \text{ Os erros do modelo apresentam variância constante} \\ H_1 : \text{ Os erros do modelo não apresentam variância constante} \end{cases}$$

A interpretação do teste se baseia no p -valor, onde:

- $p > 0,05 \Rightarrow$ Não rejeitamos H_0 , ou seja, não há evidências de heterocedasticidade.
- $p \leq 0,05 \Rightarrow$ Rejeitamos H_0 , indicando heterocedasticidade significativa.

Tabela 22: Teste de homoscedasticidade dos erros do modelo

Modelo	p -valor
Transformação logarítmica com $\log(\text{Médicos por } 100.000 \text{ hab.})$	0.8468

Como apresentado pela Tabela 22 o p -valor do teste de hipótese associado ao modelos é de 0.8468. Como o nível de significância estabelecido para o teste foi de 5%, e o p -valor do teste é superior a este valor, não existem evidências para rejeitar a hipótese nula. Dessa forma, confirmando a suspeita observada na Figura 64, o modelo é homoscedástico.

4.3 Não-correlação dos erros

Neste tópico será avaliada a suposição de não-correlação dos erros dos modelos de regressão ajustados que aceitaram os pressupostos de normalidade e homocedasticidade dos erros. Para isso, será utilizado o teste de Durbin-Watson e gráficos dos resíduos studentizados vs. índices.

4.3.1 Taxa de fecundidade vs. Nascimentos assistidos por profissionais de saúde qualificados

As figuras e os testes a seguir têm o objetivo de avaliar a não-correlação dos erros e fazem referência aos modelos transformados ajustados para a taxa de fecundidade vs. nascimentos assistidos por profissionais de saúde qualificados.

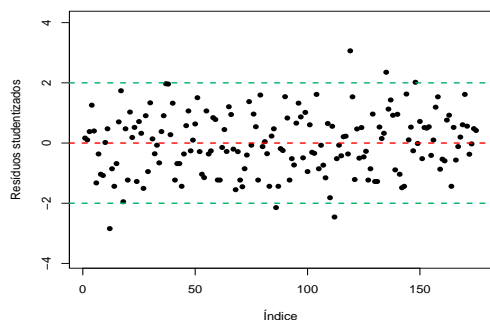


Figura 65: Modelo com transformação Box-Cox

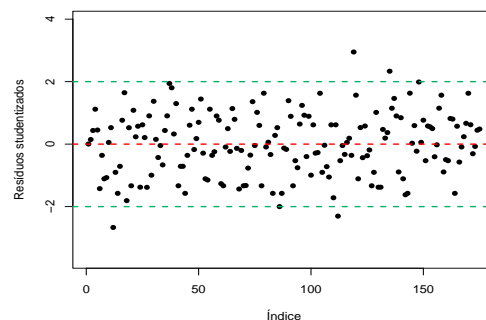


Figura 66: Modelo com transformação logarítmica

Observando as Figuras 65 e 66, não é possível perceber um padrão na distribuição dos pontos em nenhum dos modelos, ou seja, os pontos estão distribuídos de maneira aleatória ao longo do eixo dos índices. Além disso, a maioria dos pontos estão em torno do zero. Dessa forma, suspeita-se de que ambos os modelos cumprem o pressuposto de não-correlação dos erros. Para verificar esta suspeita, foi realizado um teste de hipóteses de Durbin-Watson. As hipóteses a serem testadas são:

$$\begin{cases} H_0 : \text{Não há autocorrelação entre os erros} \\ H_1 : \text{Há autocorrelação entre os erros} \end{cases}$$

A interpretação do teste se baseia no p -valor, onde:

- $p > 0,05 \Rightarrow$ Não rejeitamos H_0 , ou seja, não há evidências de autocorrelação dos erros.
- $p \leq 0,05 \Rightarrow$ Rejeitamos H_0 , indicando presença significativa de autocorrelação dos erros.

Tabela 23: Teste de não-correlação dos erros dos modelos

Modelo	p -valor
Transformação Box-Cox	0.362
Transformação logarítmica	0.325

A Tabela 23 apresenta os p -valores dos testes de hipóteses associados aos modelos com transformação de Box-Cox e com transformação logarítmica. Como o nível de significância estabelecido para os testes foi de 5%, e os p -valores foram 0.362 e 0.325, respectivamente, não existem evidências para rejeitar as hipóteses nulas de nenhum dos testes, uma vez que são maiores que o nível de significância do teste. Sendo assim, confirmando as suspeitas observadas nas Figuras 62 e 63, os erros dos modelos não são autocorrelacionados, sendo aceito o pressuposto de não-correlação para ambos os modelos.

4.3.2 Taxa de fecundidade vs. Médicos por 100.000 habitantes

A figura e o teste a seguir têm o objetivo de avaliar a não-correlação dos erros e fazem referência ao modelo transformado ajustado para a taxa de fecundidade vs. médicos por 100.000 habitantes.

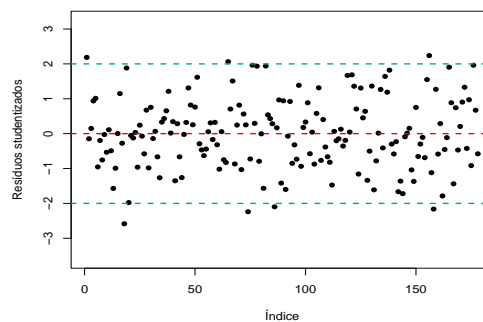


Figura 67: Modelo com transformação logarítmica e com $\log(\text{Médicos por } 100.000 \text{ hab.})$

Analisando a Figura 67, pode-se perceber que os pontos estão distribuídos de maneira aleatória ao longo do eixo dos índices. Além disso, a maioria dos pontos estão em torno do zero. Desse modo, existe uma suspeita de que o modelo não apresenta autocorrelação entre os erros. Para verificar esta suspeita, foi realizado um teste de hipóteses de Durbin-Watson. As hipóteses a serem testadas são:

$$\begin{cases} H_0 : \text{ Não há autocorrelação entre os erros} \\ H_1 : \text{ Há autocorrelação entre os erros} \end{cases}$$

A interpretação do teste se baseia no p -valor, onde:

- $p > 0,05 \Rightarrow$ Não rejeitamos H_0 , ou seja, não há evidências de autocorrelação dos erros.
- $p \leq 0,05 \Rightarrow$ Rejeitamos H_0 , indicando presença significativa de autocorrelação dos erros.

Tabela 24: Teste de não-correlação dos erros do modelo

Modelo	p -valor
Transformação logarítmica com $\log(\text{Médicos por } 100.000 \text{ hab.})$	0.601

A Tabela 24 apresenta o p -valor do teste de hipótese, que foi de 0.601. Como o nível de significância estabelecido para os testes foi de 5%, não existem evidências para rejeitar a hipótese nula do teste. Sendo assim, confirmando as suspeitas observadas na Figura 67, os erros do modelo não são autocorrelacionados, sendo aceito o pressuposto de não-correlação.

4.4 Alavancagem

Neste tópico será avaliada a presença de pontos de alavancagem nos modelos de regressão que atenderam aos pressupostos de normalidade, homocedasticidade e não correlação dos erros. Para tal, será utilizado o gráfico da medida h , que é a medida de alavancagem, versus os índices. Os limites usados de suporte para essa avaliação foram $4/n$, onde os pontos acima desse limite representam uma alavancagem moderada, e $6/n$, em que os pontos acima desse limite representam uma alavancagem alta.

4.4.1 Taxa de fecundidade vs. Nascimentos assistidos por profissionais de saúde qualificados

As figuras a seguir visam identificar os possíveis pontos de alavancagem dos modelos transformados ajustados para a taxa de fecundidade vs. nascimentos assistidos por profissionais de saúde qualificados.

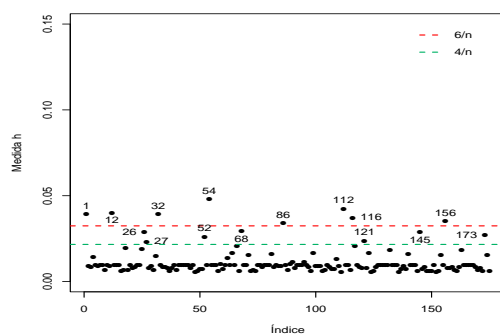


Figura 68: Modelo com transformação Box-Cox

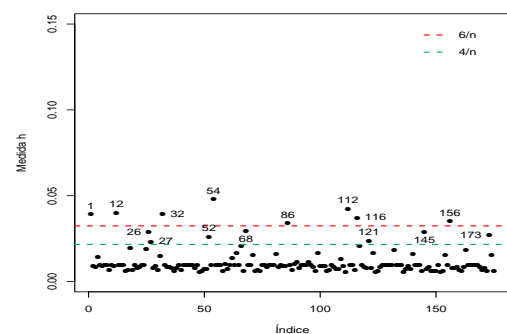


Figura 69: Modelo com transformação logarítmica

A Figura 68 ilustra as medidas de alavancagem do modelo com transformação de Box-Cox. Pode-se identificar 8 pontos acima do limite da linha vermelha, significando que podem apresentar um peso desproporcional em relação ao seu valor ajustado. Esses pontos foram: 1, 12, 32, 54, 86, 112, 116 e 156. Alguns desses pontos serão investigados posteriormente. Vale ressaltar que também foram identificados 7 pontos acima da linha verde, que podem apresentar uma alavancagem moderada. Já a Figura 69 mostra as medidas de alavancagem do modelo com transformação logarítmica. Neste modelo, identifica-se os mesmos pontos acima da linha vermelha e da linha verde, pontos estes que podem apresentar um peso desproporcional em relação ao seu valor ajustado. Assim como citado anteriormente, estes pontos também serão investigados.

4.4.2 Taxa de fecundidade vs. Médicos por 100.000 habitantes

A figura a seguir visa identificar os possíveis pontos de alavancagem do modelo transformado ajustado para a taxa de fecundidade vs. médicos por 100.000 habitantes.

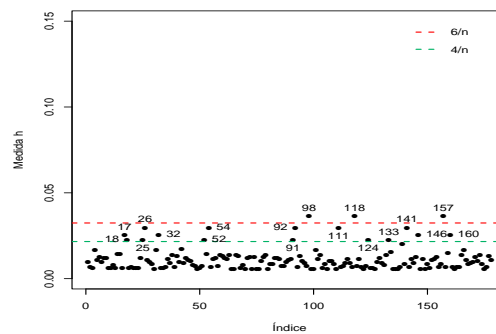


Figura 70: Modelo com transformação logarítmica e com $\log(\text{Médicos por } 100.000 \text{ hab.})$

A Figura 70 mostra as medidas de alavancagem do modelo com transformação logarítmica e com logaritmo na variável médicos por 100.000 habitantes. Pode-se identificar 3 pontos acima do limite da linha vermelha, significando que podem apresentar um peso desproporcional em relação ao seu valor ajustado. Esses pontos foram: 98, 118 e 157. Alguns desses pontos serão investigados posteriormente. É importante ressaltar que também foram identificados 15 pontos acima da linha verde, que podem apresentar uma alavancagem moderada.

4.5 Influência

Nesta seção será avaliada a presença de observações influentes nos modelos de regressão que atenderam aos pressupostos de normalidade, homocedasticidade e não correlação dos erros. Para tal, serão utilizadas medidas de influência como a distância de Cook, DFBE-TAS (para o intercepto e o coeficiente angular) e de DFFITS. Essas medidas viabilizam a identificação de observações que podem exercer um impacto na estimação dos coeficientes ou nos valores ajustados do modelo.

4.5.1 Taxa de fecundidade vs. Nascimentos assistidos por profissionais de saúde qualificados

As figuras a seguir têm o objetivo de identificar os possíveis pontos influentes nos modelos transformados ajustados para a taxa de fecundidade vs. nascimentos assistidos por profissionais de saúde qualificados.

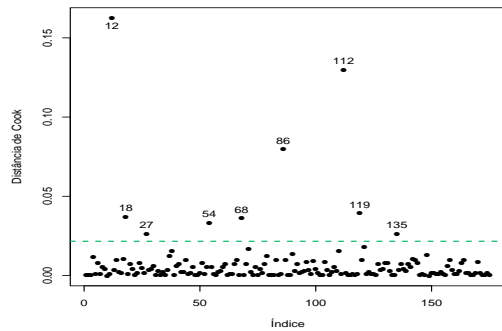


Figura 71: Modelo com transformação Box-Cox

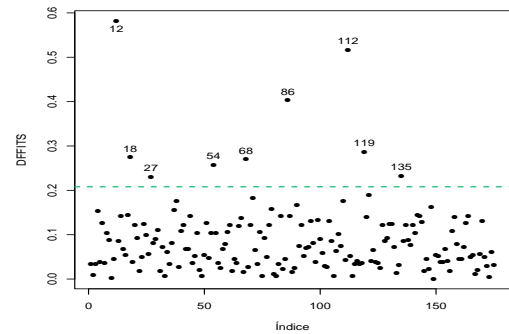


Figura 72: Modelo com transformação Box-Cox

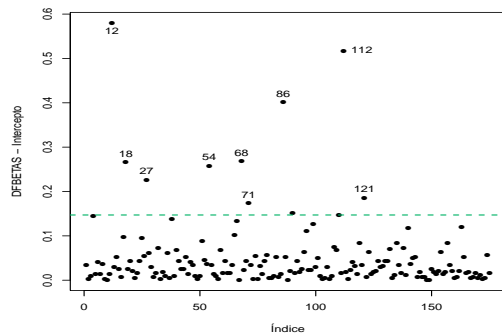


Figura 73: Modelo com transformação Box-Cox

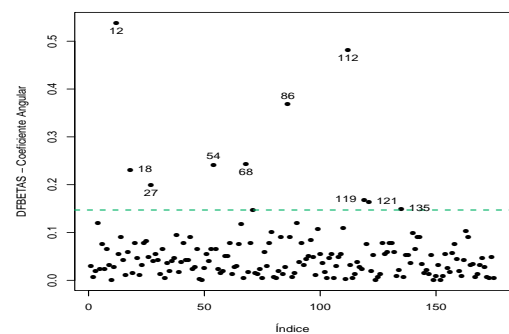


Figura 74: Modelo com transformação Box-Cox

Em relação ao modelo com transformação de Box-Cox, as medidas de influência Distância de Cook e DFFITS ilustradas pelas Figuras 71 e 72 mostram 9 possíveis pontos de influência, eles sendo: 12, 18, 27, 54, 68, 86, 112, 119 e 135. Já na medida DFBETAS - intercepto e DFBETAS - coeficiente angular, ilustradas pelas Figuras 73 e 74 mostram 11 possíveis pontos de influência, eles sendo: 12, 18, 27, 54, 68, 71, 86, 112, 119, 121 e 135. Alguns desses pontos serão investigados posteriormente.

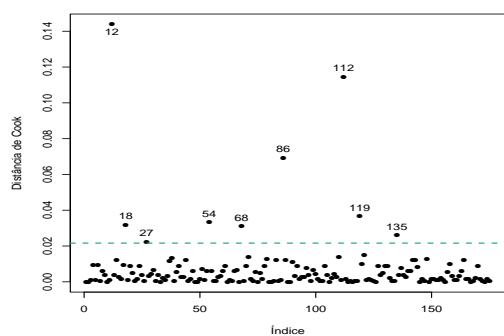


Figura 75: Modelo com transformação logarítmica

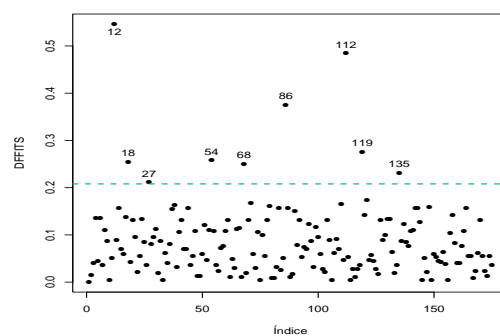


Figura 76: Modelo com transformação logarítmica

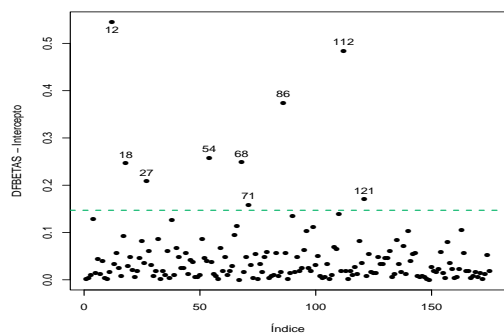


Figura 77: Modelo com transformação logarítmica

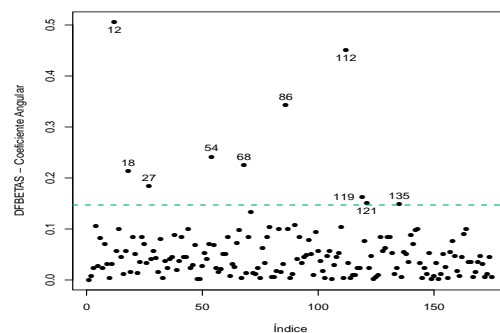


Figura 78: Modelo com transformação logarítmica

Já em relação ao modelo com transformação logarítmica, as medidas de influência Distância de Cook e DFFITS ilustradas pelas Figuras 75 e 76 mostram 9 possíveis pontos de influência, eles sendo: 12, 18, 27, 54, 68, 86, 112, 119, 135. Para a medida DFBETAS - intercepto foram identificados os pontos 12, 18, 27, 54, 68, 71, 86, 112 e 121 como possíveis pontos influentes. Já para a medida DFBETAS - coeficiente angular os pontos destacados foram os pontos 12, 18, 27, 54, 68, 86, 112, 119, 121 e 135. Alguns desses pontos serão investigados posteriormente.

4.5.2 Taxa de fecundidade vs. Médicos por 100.000 habitantes

As figuras a seguir têm o objetivo de identificar os possíveis pontos influentes do modelo transformado ajustado para a taxa de fecundidade vs. médicos por 100.000 habitantes.

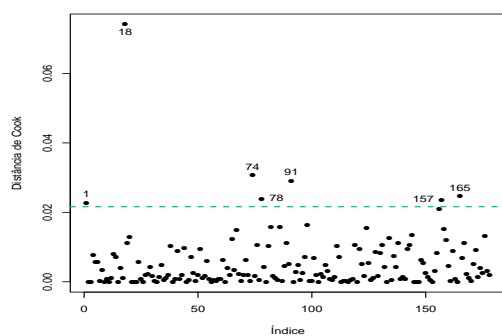


Figura 79: Modelo com transformação logarítmica e com $\log(\text{Médicos por } 100.000 \text{ hab.})$

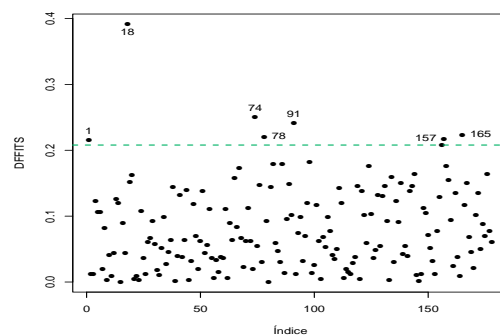


Figura 80: Modelo com transformação logarítmica e com $\log(\text{Médicos por } 100.000 \text{ hab.})$

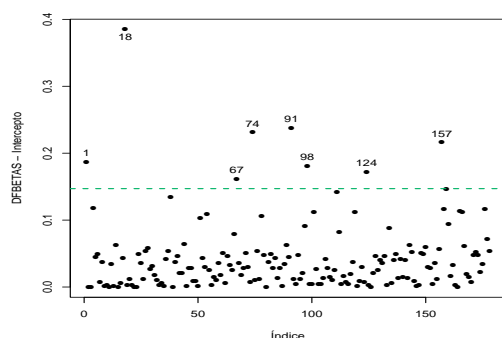


Figura 81: Modelo com transformação logarítmica e com $\log(\text{Médicos por } 100.000 \text{ hab.})$

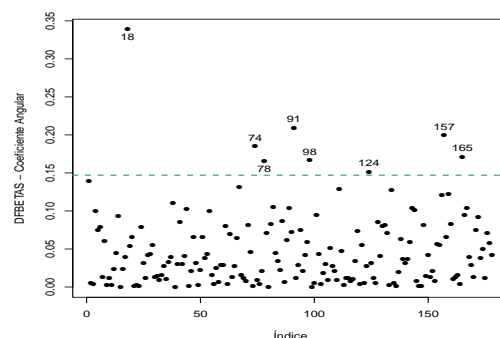


Figura 82: Modelo com transformação logarítmica e com $\log(\text{Médicos por } 100.000 \text{ hab.})$

As medidas de influência Distância de Cook e DFFITS representadas pelas Figuras 79 e 80 mostram 7 possíveis pontos de influência, eles sendo: 1, 18, 74, 78, 91, 157 e 165. Para a medida DFBETAS - intercepto ilustrada pela Figura 81 foram identificados 8 possíveis pontos de influência, eles sendo: 1, 18, 67, 74, 91, 98, 124 e 157. Já para a medida DFBETAS - coeficiente angular os possíveis pontos de influência identificados foram: 18, 74, 78, 91, 98, 124, 157 e 165, como mostra a Figura 82. Como citado nas análises anteriores, alguns destes pontos serão investigados posteriormente.

5 Investigação de pontos atípicos

Nesta seção serão investigados alguns dos pontos destacados nas seções anteriores. Essa investigação será realizada através da retirada desses pontos de maneira individual e conjunta para avaliarmos o impacto nas estatísticas do modelo e se existe ou não mudanças inferenciais.

5.1 Taxa de fecundidade vs. Nascimentos assistidos por profissionais de saúde qualificados

Os pontos com mais destaque nas medidas de influência e alavancagem no modelo com transformação de Box-Cox foram os pontos 12, 112 e 86. Segue a tabela com as mudanças relativas nos coeficientes do modelo ao realizar cada retirada.

Tabela 25: Estimativas do modelo com transformação de Box-Cox retirando os pontos atípicos

Pontos	β_0	Mudança em β_0	β_1	Mudança em β_1
Com todos os pontos	2.681785	0%	-0.019088	0%
Retirando o ponto 12	2.683182	0.05%	-0.019120	0.17%
Retirando o ponto 112	2.681586	-0.01%	-0.019083	-0.03%
Retirando o ponto 86	2.681405	-0.01%	-0.019080	-0.04%
Retirando os pontos 12 e 112	2.682992	0.05%	-0.019116	0.15%
Retirando os pontos 12 e 86	2.682813	0.04%	-0.019112	0.13%
Retirando os pontos 112 e 86	2.681201	-0.02%	-0.019075	-0.07%
Retirando os pontos 12, 112, 86	2.682618	0.03%	-0.019107	0.1%

Observando os resultados obtidos pela investigação dos pontos atípicos do modelo com transformação de Box-Cox, expressos na Tabela 25, pode-se dizer que não houve nenhuma mudança considerável nos coeficientes β_0 e β_1 , sendo todos menores que 0.20%. Além disso, também não ocorreu nenhuma mudança inferencial em relação aos p-valores dos modelos, tornando este modelo um forte candidato para explicar a variável resposta.

No modelo com transformação logarítmica, assim como no modelo anterior, os pontos nas medidas de influência e alavancagem foram os pontos 12, 112 e 86. Segue a tabela com as mudanças relativas nos coeficientes do modelo ao realizar cada retirada.

Tabela 26: Estimativas do modelo com transformação de logarítmica retirando os pontos atípicos

Pontos	β_0	Mudança em β_0	β_1	Mudança em β_1
Com todos os pontos	2.2262317	0%	-0.0151758	0%
Retirando o ponto 12	2.2275052	0.06%	-0.0152052	0.19%
Retirando o ponto 112	2.226152	0%	-0.0151743	-0.01%
Retirando o ponto 86	2.2259915	-0.01%	-0.0151704	-0.04%
Retirando os pontos 12 e 112	2.2274348	0.05%	-0.0152035	0.18%
Retirando os pontos 12 e 86	2.227276	0.05%	-0.015200	0.16%
Retirando os pontos 112 e 86	2.2259087	-0.01%	-0.0151684	-0.05%
Retirando os pontos 12, 112, 86	2.2272029	0.04%	-0.0151984	0.15%

Assim como na análise feita para o modelo anterior, os resultados obtidos na investigação dos pontos atípicos do modelo com transformação logarítmica estão expressos na Tabela 26 e não ilustram mudanças significativas nos coeficientes β_0 e β_1 , sendo elas positivas ou negativas, de maneira que são todas menores que 0.20%. Ademais, também não

ocorreram mudanças inferenciais em relação aos p-valores dos modelos, fator que torna este modelo um bom candidato para explicar a variável resposta.

5.2 Taxa de fecundidade vs. Médicos por 100.000 habitantes

Os pontos com mais destaque nas medidas de influência e alavancagem no modelo em questão foram os pontos 18, 74, 91 e 157. Segue a tabela com as mudanças relativas nos coeficientes do modelo ao realizar cada retirada.

Tabela 27: Estimativas do modelo com transformação logarítmica e com $\log(\text{Médicos por } 100.000 \text{ hab.})$ retirando os pontos atípicos

Pontos	β_0	Mudança em β_0	β_1	Mudança em β_1
Com todos os pontos	2.25916	0%	-0.29310	0%
Retirando o ponto 18	2.26196	0.12%	-0.29365	0.19%
Retirando o ponto 74	2.25596	-0.14%	-0.29194	-0.39%
Retirando o ponto 91	2.26135	0.1%	-0.29398	0.3%
Retirando o ponto 157	2.25538	-0.17%	-0.29249	-0.21%
Retirando os pontos 18 e 74	2.25871	-0.02%	-0.29249	-0.21%
Retirando os pontos 18 e 91	2.26418	0.22%	-0.29455	0.5%
Retirando os pontos 18 e 157	2.25810	-0.05%	-0.29303	-0.02%
Retirando os pontos 74 e 91	2.25815	-0.04%	-0.29283	-0.09%
Retirando os pontos 74 e 157	2.25219	-0.31%	-0.29134	-0.6%
Retirando os pontos 91 e 157	2.25815	-0.04%	-0.29283	-0.09%
Retirando os pontos 18, 74, 91	2.26093	0.08%	-0.29339	0.1%
Retirando os pontos 18, 74, 157	2.25485	-0.19%	-0.29187	-0.42%
Retirando os pontos 18, 91, 157	2.26031	0.05%	-0.29392	0.28%
Retirando os pontos 74, 91, 157	2.25438	-0.21%	-0.29223	-0.3%
Retirando os pontos 18, 74, 91 e 157	2.25707	-0.09%	-0.29277	-0.11%

Observando os resultados obtidos pela investigação dos pontos atípicos observados no modelo com transformação logarítmica e com logaritmo na variável médicos por 100.000 habitantes, expressos na Tabela 27, pode-se dizer que não houve nenhuma mudança significativa nos coeficientes β_0 e β_1 , tanto positivas quanto negativas, sendo todas menores que 0.5%. Além disso, também houve mudanças inferenciais em relação aos p-valores dos modelos, tornando este modelo possível candidato para explicar a variável resposta.

6 Interpretação do modelo selecionado

Nesta seção, o modelo de regressão linear simples escolhido será interpretado de maneira mais profunda. Além disso, uma tabela com todos os modelos ajustados e os seus resultados mediante as suposições necessárias para uma análise de regressão linear também será apresentada.

Tabela 28: Resumo dos modelos de regressão linear simples propostos

Número do modelo	Variável resposta	Covariável	Suposições	Pontos atípicos
1	Taxa de fecundidade	Expectativa de vida	Desvio da homoscedasticidade	-
2	Taxa de fecundidade	Taxa de analfabetismo adulto	Desvio da normalidade	-
3	$\sqrt{\text{Taxa de fecundidade}}$	Taxa de analfabetismo adulto	Desvio da normalidade	-
4	Taxa de fecundidade	$\log(\text{Gasto com saúde per capita em 2004})$	Desvio da homoscedasticidade	-
5	$\log(\text{Taxa de fecundidade})$	$\log(\text{Gasto com saúde per capita em 2004})$	Desvio da normalidade	-
6	Taxa de fecundidade	Nascimentos assistidos por profissionais	Desvio da normalidade	-
7	$(\text{Taxa de fecundidade})^{0.2} - 1/0.2$	Nascimentos assistidos por profissionais	Aceita todas as suposições	12, 112 e 86
8	$\log(\text{Taxa de fecundidade})$	Nascimentos assistidos por profissionais	Aceita todas as suposições	12, 112 e 86
9	Taxa de fecundidade	$\log(\text{médicos por 100.000 habitantes})$	Desvio da normalidade	-
10	$\log(\text{Taxa de fecundidade})$	$\log(\text{médicos por 100.000 habitantes})$	Aceita todas as suposições	18, 74, 91 e 157
11	Taxa de fecundidade	$\log(\text{Pesquisadores em Pesquisa e Desenvolvimento})$	Desvio da normalidade	-
12	$\log(\text{Taxa de fecundidade})$	$\log(\text{Pesquisadores em Pesquisa e Desenvolvimento})$	Desvio da normalidade	-

De acordo com a análise diagnóstica realizada, o melhor modelo de regressão simples que foi ajustado com o objetivo de explicar a variável taxa de fecundidade foi o **Modelo 10 - $\log(\text{Taxa de fecundidade})$ vs. $\log(\text{médicos por 100.000 habitantes})$** , pois este apresenta o maior coeficiente de determinação dentre os modelos que demonstraram estar adequados com relação às suposições necessárias para suas devidas inferências, sendo ele $R^2 = 0.707$, ou seja, o modelo explica mais de 70% da variabilidade do logaritmo da resposta. Além disso, o erro padrão residual é de 0.287, o menor entre todos os modelos.

Tabela 29: Ajuste do modelo com transformação logarítmica e $\log(\text{Médicos por 100.000 hab.})$

Estatística	Valor
Intercepto (β_0)	2.259
Erro Padrão do Intercepto	0.064
Coeficiente (β_1)	-0.293
Erro Padrão do Coeficiente	0.014
R^2	0.707
R^2 Ajustado	0.705
Erro Padrão Residual	0.287
p -valor do Modelo	2.2×10^{-16}

Interpretando o modelo ajustado com mais detalhes, tem-se que $\beta_0 = 2.259$, indicando que a média do logaritmo da taxa de fecundidade quando fixamos o valor do logaritmo da covariável médicos por 100.000 habitantes à zero, mas levando em consideração que $\log(0) = 1$, pode-se dizer que é igual a fixarmos a 1. Esta interpretação de β_0 pode ser

rara, entretanto, possível. Quanto a interpretação do β_1 , que é igual a 0.293, é a variação na média do logaritmo da taxa de fecundidade quando se acrescenta uma unidade no logaritmo de médicos por 100.000 habitantes.

Entretanto, como foi aplicada uma transformação na variável resposta para que o modelo fosse melhor ajustado, os coeficientes não correspondem mais à escala original, mas sim à escala logarítmica. Apesar de dificultar um pouco a interpretação geral do modelo, ela ainda é possível, basta aplicar o exponencial nas estimativas do modelo. Sendo assim, tem-se que $\exp(\beta_0) = 9.574$ e $\exp(\beta_1) = 1.34$, ou seja, 9.574 é a média da resposta taxa de fecundidade quando fixamos o valor do logaritmo da covariável médicos por 100.000 habitantes à 1, e 1.34 signifique que, ao aumentar em um ponto o logaritmo de médicos por 100.000, a média da taxa de fecundidade aumenta em 34%.

Ademais, também foi realizado o cálculo do intervalo de confiança de 95% para os coeficientes do modelo selecionado.

Tabela 30: Intervalos de confiança dos coeficientes do modelo selecionado

Coeficientes	Limite Inferior	Limite Superior
Intercepto	2.133	2.385
log(médicos por 100.000 habitantes)	-0.321	-0.265