



UNIVERSIDADE FEDERAL DA BAHIA
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DEPARTAMENTO DE ESTATÍSTICA

CAMILLE MENEZES PEREIRA DOS SANTOS
MICHEL MILER ROCHA DOS SANTOS

ANÁLISE DA RELAÇÃO ENTRE A RENDA MENSAL E
VARIÁVEIS SOCIOECONÔMICAS, DEMOGRÁFICAS E
CLIMÁTICAS EM VÁRIOS PAÍSES: UM ESTUDO DE
REGRESSÃO LINEAR

Salvador
2023

**CAMILLE MENEZES PEREIRA DOS SANTOS
MICHEL MILER ROCHA DOS SANTOS**

**ANÁLISE DA RELAÇÃO ENTRE A RENDA MENSAL E
VARIÁVEIS SOCIOECONÔMICAS, DEMOGRÁFICAS E
CLIMÁTICAS EM VÁRIOS PAÍSES: UM ESTUDO DE
REGRESSÃO LINEAR**

Atividade de laboratório apresentada ao Instituto de Matemática e Estatística da Universidade Federal da Bahia como parte das exigências da disciplina Análise de Regressão ministrada pela professora Dra. Edleide de Brito.

Salvador
2023

SUMÁRIO

1	INTRODUÇÃO	1
2	MATERIAIS E MÉTODOS	2
2.1	Descrição dos dados	2
3	RESULTADOS	3
3.1	Análise Descritiva	3
3.2	Modelos de regressão linear	5
3.3	Análise dos resíduos	8
4	CONCLUSÃO	10

1 Introdução

Este trabalho tem como objetivo analisar a relação entre a renda mensal e a expectativa de vida masculina em diferentes continentes. Utilizando dados de países ao redor do mundo em 2019, foram desenvolvidos modelos de regressão linear para investigar como essas variáveis estão relacionadas.

A análise descritiva inicial revelou diferenças significativas entre os continentes em termos de renda e expectativa de vida masculina. Verificou-se que a renda mensal apresenta uma relação exponencial com a expectativa de vida masculina, sugerindo que um aumento na expectativa de vida está associado a um aumento não linear na renda mensal. Para melhor explorar essa relação, foi aplicada a transformação logarítmica na renda, que ajudou a linearizar essa relação e a capturar de forma mais adequada a associação entre as variáveis.

Com base nessas observações, foram realizadas análises de regressão linear, considerando diferentes modelos. O modelo nulo, que não inclui nenhuma variável explicativa, o qual a estimativa é a média amostral da variável resposta. Os modelos de retas coincidentes, retas paralelas e retas concorrentes foram adotados para examinar tanto os efeitos individuais da expectativa de vida masculina na renda quanto os efeitos do continente na relação entre essas variáveis. Para escolher entre os modelos citados qual foi o modelo mais adequado, será realizado o teste F parcial, que é capaz de analisar a igualdade de modelos "encaixados", que é o caso desses modelos.

Além disso, será realizada uma análise dos resíduos para verificar se o modelo escolhido atende aos pressupostos da regressão linear. Os resíduos representam as diferenças entre os valores observados e os valores estimados pelo modelo e são utilizados para avaliar a qualidade do ajuste. Eventuais pontos discrepantes ou desvios dos pressupostos serão cuidadosamente considerados, a fim de garantir uma análise mais precisa e confiável dos resultados.

2 Materiais e métodos

2.1 Descrição dos dados

A base de dados contém informações socioeconômicas, demográficas e climáticas de 82 países, disponíveis no site <https://www.dadosmundiais.com/>. Os dados foram obtidos por meio da combinação de bases de dados contendo informações sobre a expectativa de vida, quociente de inteligência e custo de vida de cada país.

As variáveis presentes no conjunto de dados são:

Variável	Descrição
País	Nome do país
<u>QI</u>	<u>Quociente de inteligência</u>
Despesas com educação	Gastos do estado com educação em dólares
Temperatura máxima diária	Temperatura máxima diária em graus celsius
Renda mensal	Renda mensal, em dolar, calculada a partir da renda nacional bruta por habitante
Expectativa de vida masculina	Expectativa de vida masculina em anos
Expectativa de vida feminina	Expectativa de vida feminina em anos
Taxa de natalidade	Taxa de nascimento por 1000 habitantes
Taxa de mortalidade	Taxa de morte por 1000 habitantes
Continente	Continente em que os países estão localizados

Tabela 2.1: Variáveis disponíveis no conjunto de dados

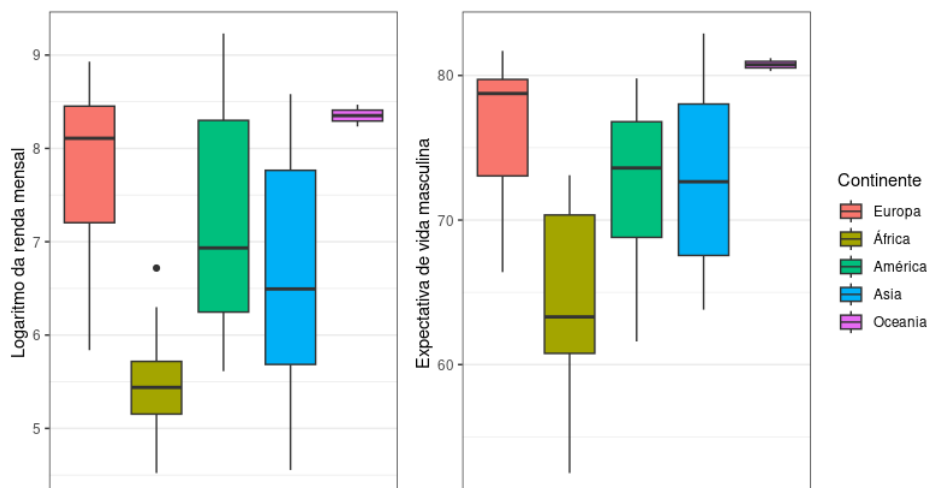
3 Resultados

3.1 Análise Descritiva

Como foi visto no Laboratório 4, onde foi construído um modelo linear simples com cada variável contínua e com a variável renda mensal como variável resposta, a expectativa de vida masculina apresenta uma relação exponencial com a variável resposta renda mensal. Por isso, a transformação logarítmica da renda mensal pode ser considerada, isso ajuda a reduzir a escala dos valores e tornar a relação entre as variáveis mais linear.

Observando a Figura 3.1, é possível notar que a Europa e a Oceania apresentam os maiores logaritmos da renda mensal e a África apresenta os menores. Mas, a Oceania tem apenas dois países: Austrália e Nova Zelândia, por isso é mais complicado analisar as características de dispersão das variáveis em relação a esse continente.

Figura 3.1: *Boxplots*, de acordo com o continente, do logaritmo da renda mensal e da expectativa de vida masculina dos países em 2019



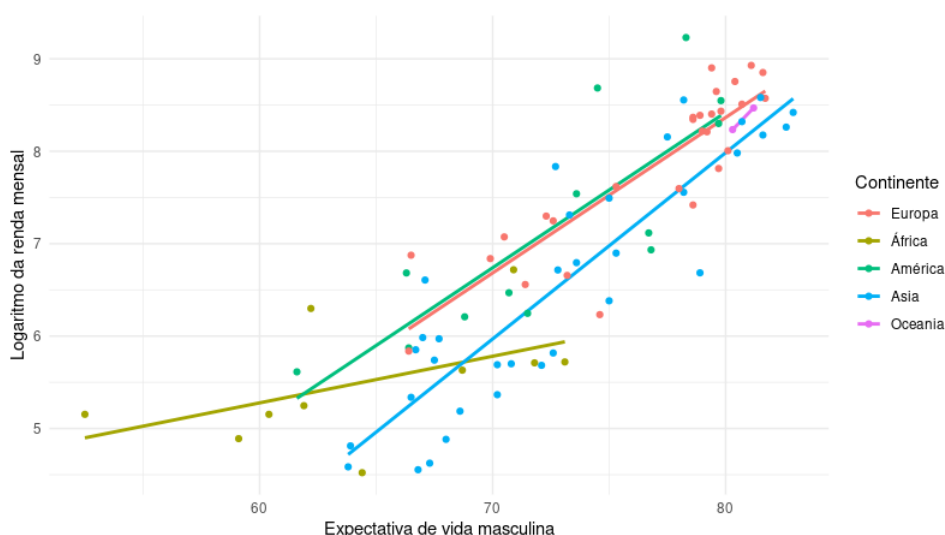
A América e a Ásia possuem uma grande variabilidade no logaritmo da renda mensal, sugerindo uma grande discrepância entre a renda dos países desses continentes. A Oceania apresenta pequena variabilidade, o que pode ser fruto do que foi dito anteriormente sobre a quantidade de países nesse continente. A Europa apresenta uma maior variabilidade

abaixo da mediana, o que pode caracterizar uma assimetria à esquerda.

A expectativa de vida masculina apresenta maiores valores na Oceania e Europa. A África apresenta os menores valores de expectativa de vida, com uma alta variabilidade entre a mediana e o terceiro quartil, indicando o que poderia ser uma assimetria à direita nessa variável. A distribuição da expectativa de vida masculina na Ásia parece ser semelhante com a da América, tendo um pouco mais de variabilidade na Ásia.

O logaritmo da renda mensal e a expectativa de vida masculina são fortemente correlacionadas, apresentam um coeficiente de correlação de 0,86. É possível notar essa relação no gráfico de dispersão entre essas variáveis, Figura 3.2.

Figura 3.2: Gráfico de densidade e curva de regressão, de acordo com o continente, dos países em 2019



Os países do continente africano, mesmo aqueles com expectativa de vida masculina mais alta, apresentam logaritmo da renda mensal menor que outros continentes (com exceção de alguns países da Ásia). O continente asiático, por sua vez, possui maior discrepância entre a renda de seus países, mas a expectativa de vida masculina deles é alta, mesmo nos países com logaritmo da renda mensal mais baixo. Além disso, o continente americano possui os mais altos logaritmos da renda mensal, seguido de perto pelo continente europeu.

É possível notar ainda que o modelo de efeitos multiplicativos com interação parece ser o mais adequado para modelar esses dados, uma vez que as curvas de regressão dos continentes são concorrentes.

3.2 Modelos de regressão linear

A Tabela 3.1 apresenta a sumarização da regressão linear sem variáveis preditoras. O coeficiente do intercepto indica a média da variável resposta, o logaritmo da renda mensal. A estimativa da renda mensal dada por esse modelo é de 1063,31 dólares, que é diferente da média da renda mensal que é 2143,34 dólares. Portanto, um modelo nulo com a transformação logarítmica já tem a capacidade reduzir o efeito produzido por valores extremos da variável renda mensal.

Tabela 3.1: Sumarização da regressão linear em que o modelo nulo é utilizado para prever o logaritmo da renda mensal

Coeficientes	Estimativa	Exp(Est.)	EP	Est. T	Valor-p
Intercepto	6,97	1063.31	0,14	49,42	<0,001

O modelo de regressão com expectativa de vida masculina como variável independente, Tabela 3.2, indica que para cada unidade de aumento na expectativa de vida masculina, a renda mensal aumenta em média $(1 - \text{Exp}(\text{est})) \cdot 100\% = 19\%$. O valor-p sugere que este coeficiente é significativo e o R^2 indica que cerca de 74% da variação na renda mensal pode ser explicada pela variação na expectativa de vida masculina.

Tabela 3.2: Sumarização da regressão linear em que a expectativa de vida masculina é utilizada para prever o logaritmo da renda mensal

Coeficientes	Estimativa	Exp(Est.)	EP	Est. T	Valor-p
Intercepto	-5,75	0,0032	0,83	-6,96	<0,001
ExpectHomens	0,17	1,1894	0,01	15,46	<0,001

A regressão linear da Tabela 3.3 é um modelo aditivo que inclui a variável preditora expectativa de vida masculina e a variável categórica continente. Para a variável categórica continente, a categoria de referência é a Europa. Dado que o modelo é aditivo, a adição de uma unidade na variável expectativa de vida masculina tem a mesma interpretação. Para cada acréscimo de uma unidade na variável expectativa de vida masculina, é esperado um aumento de 0,16 no logaritmo da renda mensal ou um aumento de 18% na renda mensal.

As interpretações dos coeficientes da variável continente são as seguintes:

- Quando o país pertence a Europa (categoria de referência), temos que todas as variáveis *dummies* dos outros continentes são iguais a zero. Então, o modelo fica $\log(Y) = -4,8 + 0,16X$ ou $Y = 0,008 \cdot \exp\{0,16X\}$.
- Quando o país pertence a África, é esperado uma redução de 26% no valor da renda mensal quando comparado aos países da Europa.

Tabela 3.3: Sumarização das regressão linear em que o modelo de efeitos aditivos com a expectativa de vida masculina e o continente são usados para predizer o logaritmo da renda mensal

Coeficientes	Estimativa	Exp(Est.)	EP	Est. T	Valor-p
Intercepto	-4,80	0,008	0,98	-4,91	<0,001
ExpectHomens	0,16	1,179	0,01	13,00	<0,001
Continente					
África	-0,30	0,741	0,28	-1,09	0,281
América	0,041	1,041	0,22	0,19	0,851
Asia	-0,63	0,531	0,17	-3,81	<0,001
Oceania	-0,12	0,883	0,46	-0,27	0,787

- Quando o país pertence a América, é esperado um aumento de 4% no valor da renda mensal quando comparado aos países da Europa.
- Quando o país pertence a Ásia, é esperado uma redução de 47% no valor da renda mensal quando comparado aos países da Europa.
- Quando o país pertence a Oceania, é esperado uma redução de 12% no valor da renda mensal quando comparado aos países da Europa.

O R^2 aumentou para 0,79, indicando que cerca de 79% da variação na renda mensal pode ser explicada pelo modelo aditivo com a expectativa de vida masculina e o continente. Realizando o teste F e comparando o modelo só com a variável explicativa expectativa de vida masculina (modelo reduzido) com o modelo aditivo com a expectativa de vida masculina e o continente (modelo completo), foi obtido um valor-p aproximado de 0,001, ou seja, rejeitando a hipótese nula de igualdade dos modelos ao nível de significância de 5%. Portanto, o impacto da inserção da variável categórica continente no modelo é significativo para explicar o logaritmo da renda mensal.

A Tabela 3.4 apresenta a sumarização da regressão linear de um modelo com interação entre a variável expectativa de vida masculina e o continente para predizer o logaritmo da renda mensal. Novamente, a Europa é a categoria de referência para variável categórica continente.

- Quando o país está na Europa, todas as variáveis *dummies* e com interação são iguais a 0 e o modelo é dado por $\log(Y) = -5,09 + 0,168X$ ou $Y = 0,0061 \cdot \exp\{0,168X\}$. Portanto, um acréscimo de uma unidade na expectativa de vida masculina aumenta renda mensal em 18% para os países europeus.
- Quando o país está na África, a reta de regressão é dada por $\log(Y) = 2,25 + 0,048X$ ou $Y = 9,49 \cdot \exp\{0,048X\}$. Portanto, um acréscimo de uma unidade na expectativa de vida masculina aumenta renda mensal em 4,91% para os países africanos.

Tabela 3.4: Sumarização das regressão linear em que o modelo de efeitos multiplicativos com a expectativa de vida masculina e o continente são usados para prever o logaritmo da renda mensal

Coeficientes	Estimativa	Exp(Est.)	EP	Est. T	Valor-p
Intercepto	-5,09	0,0061	1,85	-2,75	0,007
ExpectHomens	0,168	1,1832	0,02	6,98	<0,001
Continente					
África	7,34	1544,2	2,63	2,79	0,007
América	0,06	1,0592	2,79	0,02	0,984
Ásia	-3,05	0,0474	2,24	-1,36	0,177
Oceania	-7,58	0,0005	72,23	-0,10	0,917
ExpectHomens:África	-0,12	0,8889	0,04	-3,13	0,002
ExpectHomens:América	-0,00002	0,99998	0,04	-0,00	1,000
ExpectHomens:Ásia	0,03	1,0339	0,03	1,13	0,264
ExpectHomens:Oceania	0,09	1,0965	0,89	0,10	0,918

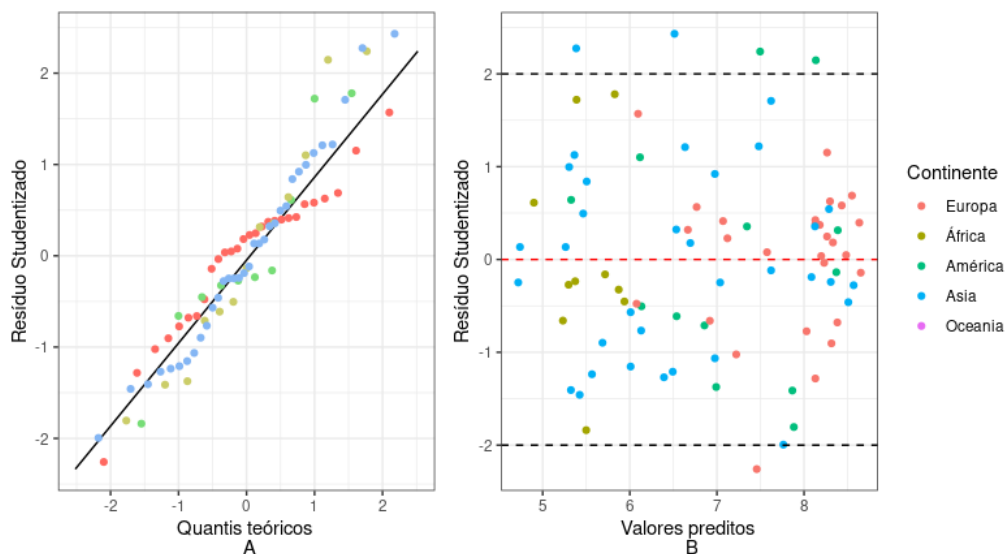
- Quando o país está na América, a reta de regressão é dada por $\log(Y) = -5,03 + 0,16798X$ ou $Y = 0,0065 \cdot \exp\{0,16798X\}$. Portanto, um acréscimo de uma unidade na expectativa de vida masculina aumenta renda mensal em 18,29% para os países africanos.
- Quando o país está na Ásia, a reta de regressão é dada por $\log = -8,14 + 0,171X$ ou $Y = 0,00029 \cdot \exp\{0,171X\}$. Portanto, um acréscimo de uma unidade na expectativa de vida masculina aumenta renda mensal em 18,64% para os países asiáticos.
- Quando o país está na Oceania, a reta de regressão é dada por $\log = -12,67 + 0,177X$ ou $Y = 0,000003 \cdot \exp\{0,177X\}$. Portanto, um acréscimo de uma unidade na expectativa de vida masculina aumenta renda mensal em 19,36% para os países da Oceania.

Para o modelo com efeitos multiplicativos, o R^2 aumentou para 0,8324, ou seja, 83,24% da variabilidade do logaritmo da renda mensal é explicada pelo modelo com interação com a expectativa de vida masculina e o continente. Para comparar o modelo aditivo com o modelo com interação, foi realizado o teste F parcial, o qual avalia a igualdade entre o modelo completo (modelo com efeitos multiplicativos) e o modelo reduzido (modelo com efeitos aditivos), e o valor-p obtido foi aproximadamente 0,001, rejeitando a hipótese de igualdade entre os modelos com 5% de significância. Portanto, a interação entre a variável expectativa de vida masculina e a variável categórica continente é relevante para explicar o logaritmo da renda mensal.

3.3 Análise dos resíduos

Em relação a [Figura 3.3 \(A\)](#), é possível notar que a dispersão dos resíduos permanece uniforme em relação aos valores preditos. O teste de *Breusch-Pagan* apontou a não rejeição da hipótese de homocedasticidade (valor-p = 0,41). Logo, o resíduos desse modelo de regressão não violam essa suposição.

Figura 3.3: Resíduos versus quantis teóricos e resíduos versus valores preditos



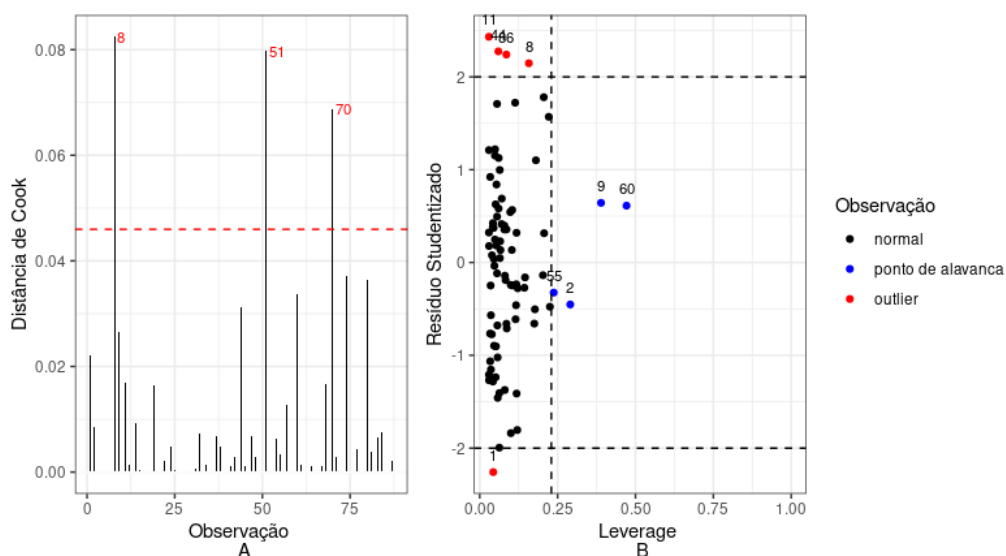
A Figura 3.4 (B), que apresenta o *qq-plot* dos resíduos Studentizados, mostra que a suposição de normalidade não é violada pelos resíduos. O teste de *Shapiro-Wilk* apresentou um valor-p de 0,65, não permitindo a rejeição da hipótese de normalidade a um nível de significância de 5%, corroborando com a análise gráfica realizada.

É possível notar ainda que os valores preditos da Europa são mais altos que dos outros continentes e os resíduos do continente asiático são mais dispersos em relação aos valores preditos. Os valores que podem ser considerados não usuais, com valor absoluto dos resíduos acima de dois, são dos continentes Europa, Ásia e América.

Já o teste de *Durbin-Watson* resultou num valor-p de aproximadamente 0,29, apontando que não há autocorrelação serial nos resíduos a um nível de 5%. Ou seja, não há a violação do pressuposto de independência.

O gráfico da distância de Cook, Figura 3.4 (A), aponta três observações como pontos influentes, isso é, são observações que afetam significativamente as estimativas do modelo de regressão. Já a Figura 3.4, apenas uma das observações que é apontada pela distância de Cook como influente é apontada como *outlier*. Nesse gráfico ainda é possível observar que há alguns *outliers* - observações que são extremas no espaço da variável resposta - no modelo e pontos de alavanca.

Figura 3.4: Gráfico da distância de Cook versus valores observados e resíduo Studentizado versus Leverage



4 Conclusão

Neste relatório, foi realizada uma análise da relação entre a renda mensal e a expectativa de vida masculina em diferentes continentes. Dado que os resultados anteriores mostraram que a renda mensal apresenta uma relação exponencial com a expectativa de vida masculina, então os modelos ajustados foram levando em consideração a transformação logarítmica para linearizar essa relação e capturar de forma mais adequada a associação entre as variáveis.

A análise descritiva revelou diferenças significativas entre os continentes em termos de renda e expectativa de vida masculina. A Europa e a Oceania apresentaram os maiores valores de renda mensal e expectativa de vida masculina, enquanto a África teve os menores valores. A América e a Ásia apresentou uma grande variabilidade na distribuição do logaritmo da renda mensal e distribuições parecidas para a expectativa de vida masculina.

Foram realizados modelos de regressão linear para investigar a relação entre a renda mensal, a expectativa de vida masculina e o continente. O modelo de efeitos multiplicativos com interação entre a expectativa de vida masculina e o continente foi considerado o mais adequado para os dados com base na análise dos testes F para modelos encaixados.

Os resultados da regressão indicaram que a expectativa de vida masculina tem um efeito positivo significativo na renda mensal. Além disso, o continente também apresentou um efeito significativo na relação entre as variáveis. A África teve uma redução significativa na renda mensal em comparação com a Europa, enquanto a América, Ásia e Oceania não apresentaram diferenças significativas.

A análise dos resíduos evidenciou que todos os pressupostos do modelo foram atendidos. Algumas observações foram identificadas como não usuais e devem ser investigadas.

Em resumo, este estudo destacou a importância da expectativa de vida masculina como um fator determinante da renda mensal em diferentes continentes. No entanto, é importante ressaltar que outras variáveis não foram consideradas, o que pode influenciar a relação entre renda e expectativa de vida.