

Uma Avaliação Empírica da Relação Entre Altura e Salários

Lucas Lourenção — 14591222

Rafael Derrico dos Santos Abreu — 14604191

1. Introdução à Questão do Estudo

1.1. A Relação Altura-Salários

Estudaremos aqui a relação entre rendimentos e altura. Esse tema já foi tratado por diversos estudos, em especial — por nos suprir com a base de dados aqui usada — o das professoras da Universidade de Princeton Anne Case e Christina Paxson, *Stature and Status: Height, Ability, and Labor Market Outcomes*, publicado em 2008 no *Journal of Political Economy*.

Ainda assim, é uma correlação que causa certa estranheza. Especialmente se se conhecem exemplos anedóticos contrários ao que indicam alguns estudos. Isto é, que trabalhadores mais bem-pagos são, em média, mais altos que seus companheiros de menor remuneração. Exploreemos, então, o porquê.

1.2. A Economia da Coisa

Case e Paxson (2008) argumentam que essa correlação, pelo menos nos países desenvolvidos, é devida a uma correlação positiva entre altura, inteligência, saúde e força física. Essas características todas promovem produtividade. Elas indicam que estatura serve de procuração para aquelas porque melhor nutrição na infância promove, além de maior estatura, maior desenvolvimento cognitivo e físico. Isto é, a maior altura não produz por si maior salário, mas indica outras características que sim.

Hübler (2006) também menciona que “a altura é função da renda, na medida em que o consumo de nutrientes, principalmente proteínas, vitaminas e minerais, e a regularidade com que são consumidos, influenciam a altura em uma determinada idade até a idade adulta. A altura média da população é um indicador da prosperidade biológica e do padrão de vida. Se os trabalhadores mais altos forem mais saudáveis, a sua produtividade deverá exceder, tudo mais constante, a dos trabalhadores mais baixos.” (p. 2).

Ele também menciona um ponto interessante: “Outra ligação entre altura e salários é induzida pela herança da altura. Um pai alto provavelmente também terá

um filho alto e uma mãe baixa provavelmente terá uma filha baixa. [...] Os pais mais instruídos e mais ricos estão mais bem informados sobre uma boa alimentação dos seus filhos e também podem comprar bens mais caros. Assim, temos de esperar que quanto mais elevada for a posição social, mais altos serão, em média, os descendentes. À medida que os pais mais instruídos souberem que os mercados pagam prêmios de altura, reforçarão os seus esforços em prol dos seus filhos.” (p. 3).

1.3. Um Modelo Preliminar

Em um primeiro momento, utilizaremos de uma regressão linear simples cujos parâmetros são obtidos pelo método *Ordinary Least Squares* (OLS). A variável dependente será o salário, e o regressor, a altura:

$$\text{Salário}_i = \beta_0 + \beta_1 \text{Altura}_i + u_i$$

Ao longo deste trabalho, este modelo será incrementado de outras variáveis e outros métodos de regressão.

2. A Base de Dados

2.1. A Base e seu Tratamento

Os dados da base foram retirados do *US National Health Interview Survey* (NHIS) de 1994 e são um subconjunto da base originalmente utilizada por Case e Paxson (2008). É relevante ressaltar que as entrevistas do NHIS são feitas não por amostragem aleatória simples, mas por técnicas de *clusters* geográficos a fim de tornar a pesquisa financeiramente viável.

Nome da Variável	Descrição
cworker	“Classe” do Trabalhador: 1 = Funcionário de empresa privada 2 = Funcionário do Governo Federal americano 3 = Funcionário de um Governo Estadual 4 = Funcionário de um Governo Local 5 = Funcionário de S/A 6 = Autônomo

earnings	Rendimentos anuais, expressos em US\$ de 2012 ¹
educ	Anos de educação formal
height	Altura, sem calçado, em polegadas
mrd	Estado civil: 1 = Casado, Cônjuge mora na mesma casa 2 = Casado, Cônjuge não mora na mesma casa 3 = Viúvo 4 = Divorciado 5 = Separado 6 = Nunca casou
occupation	Ocupação, em 15 categorias: 1 = Executivo/Gerente 2 = Profissional 3 = Técnico 4 = Vendas 5 = Administração 6 = Serviço doméstico 7 = Serviço de proteção 8 = Outro tipo de serviço 9 = Agricultura 10 = Mecânica 11 = Construção/Mineração 12 = Produção de precisão 13 = Operador de máquina 14 = Transporte 15 = Trabalhador manual
race	Raça/etnia: 1 = Branco não hispânico 2 = Negro não hispânico 3 = Hispânico 4 = Outro

¹ Na pesquisa original, os rendimentos são relatados em 23 faixas. Para cada uma dessas faixas, as professoras Case e Paxson estimaram um valor de rendimento médio com base nas informações da população atual, e esses valores médios foram atribuídos a todos os trabalhadores com rendimentos na faixa correspondente. Os valores de 1994 foram convertidos para US\$ de 2012 utilizando o índice de preços ao consumidor.

region	Região dos EUA: 1 = Nordeste 2 = Centro-Oeste 3 = Sul 4 = Oeste ²
sex	Sexo: 0 = Feminino 1 = Masculino
weight	Peso, sem calçado, em libras

Vamos também converter os dados referentes a altura e peso para o sistema métrico. De polegadas e libras para centímetros e quilogramas, respectivamente.

2.2. Análise Exploratória

Dentre as 17870 observações da base, temos 6 variáveis categóricas e 4 variáveis contínuas. Para as variáveis categóricas mais relevantes para este estudo, elaboramos gráficos com funções nativas do R e também funções da biblioteca *ggplot2* que mostram como as observações estão distribuídas:

² Segue a distribuição dos estados em cada região, conforme o Censo americano:

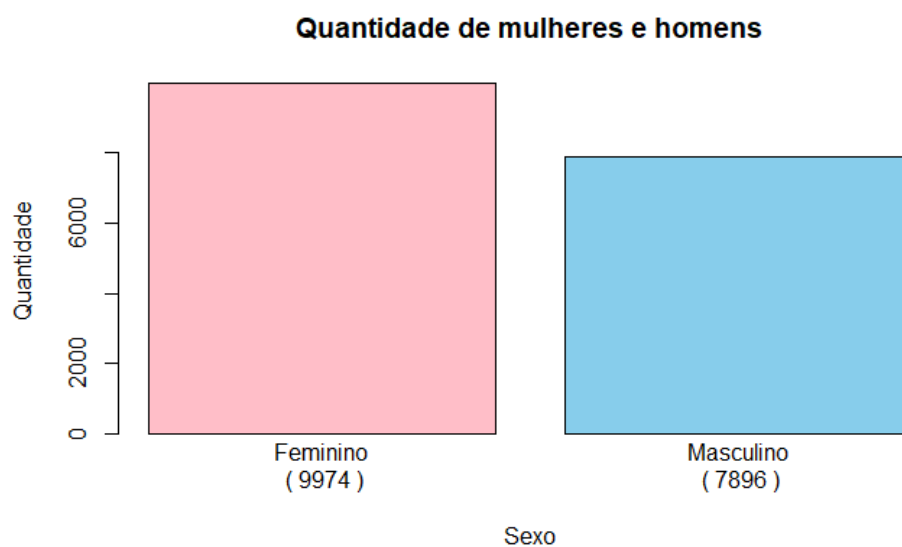
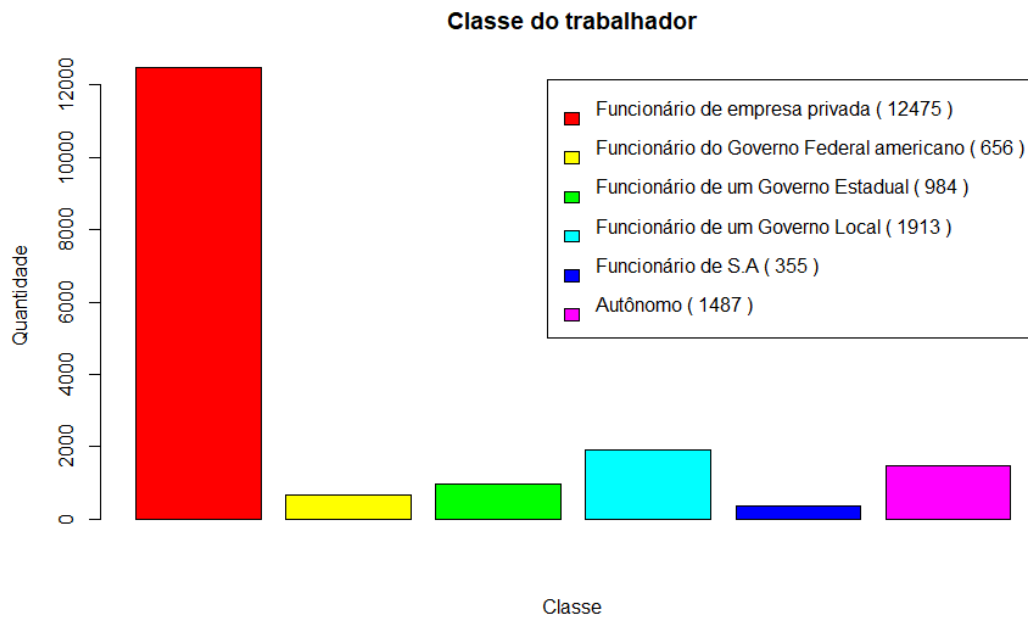
Nordeste: Connecticut, Maine, Massachusetts, Nova Hampshire, Rhode Island, Vermont, Nova Jérсия, Nova Iorque e Pensilvânia.

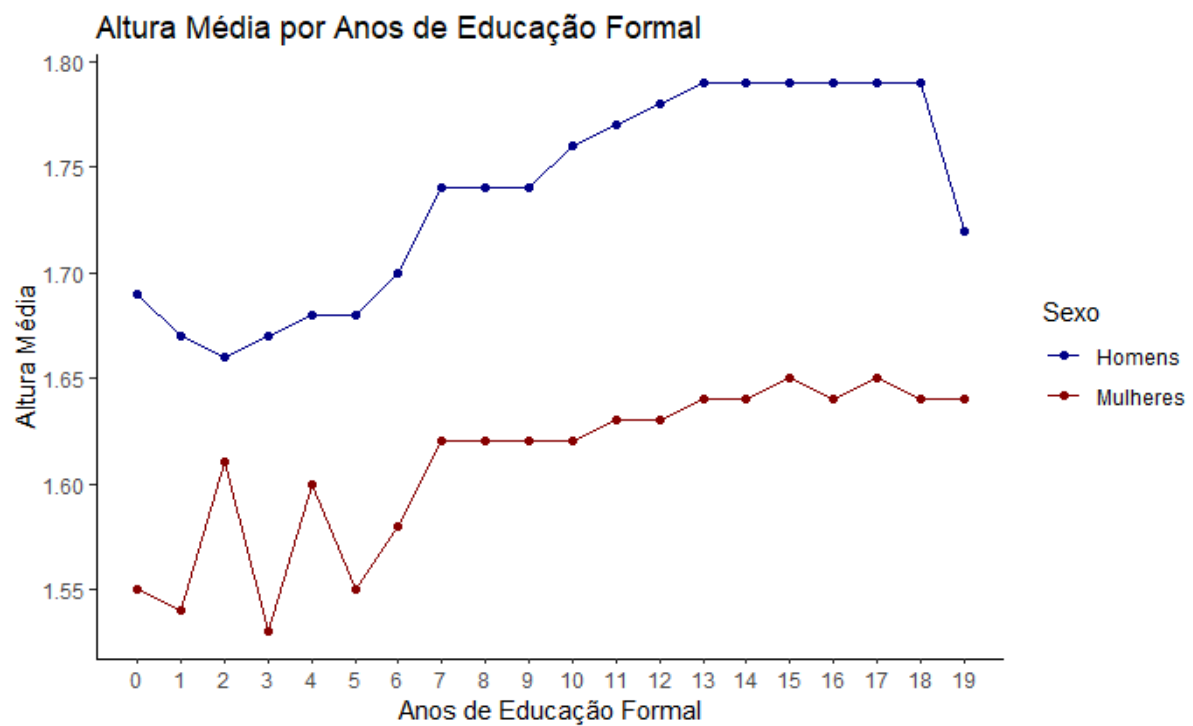
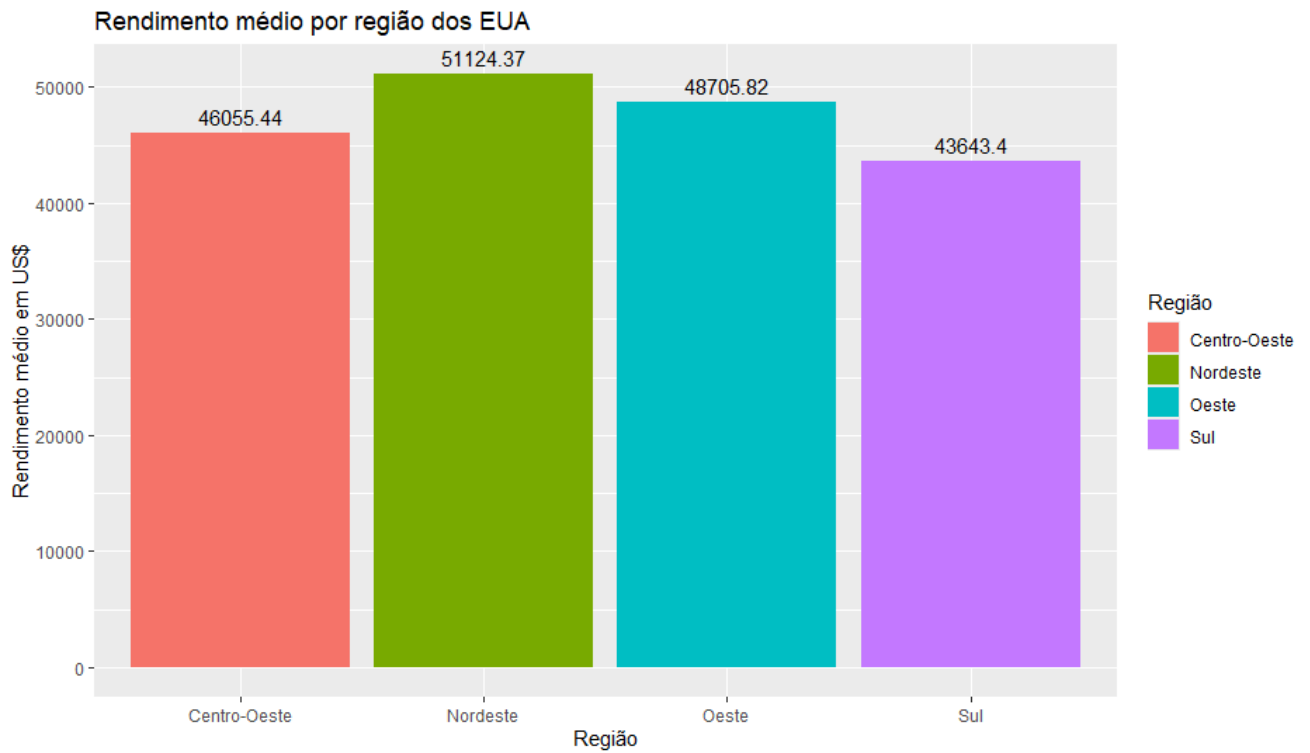
Centro-Oeste: Illinois, Indiana, Michigan, Ohio, Wisconsin, Iowa, Kansas, Minnesota, Missouri, Nebraska, Dakota do Norte e Dakota do Sul.

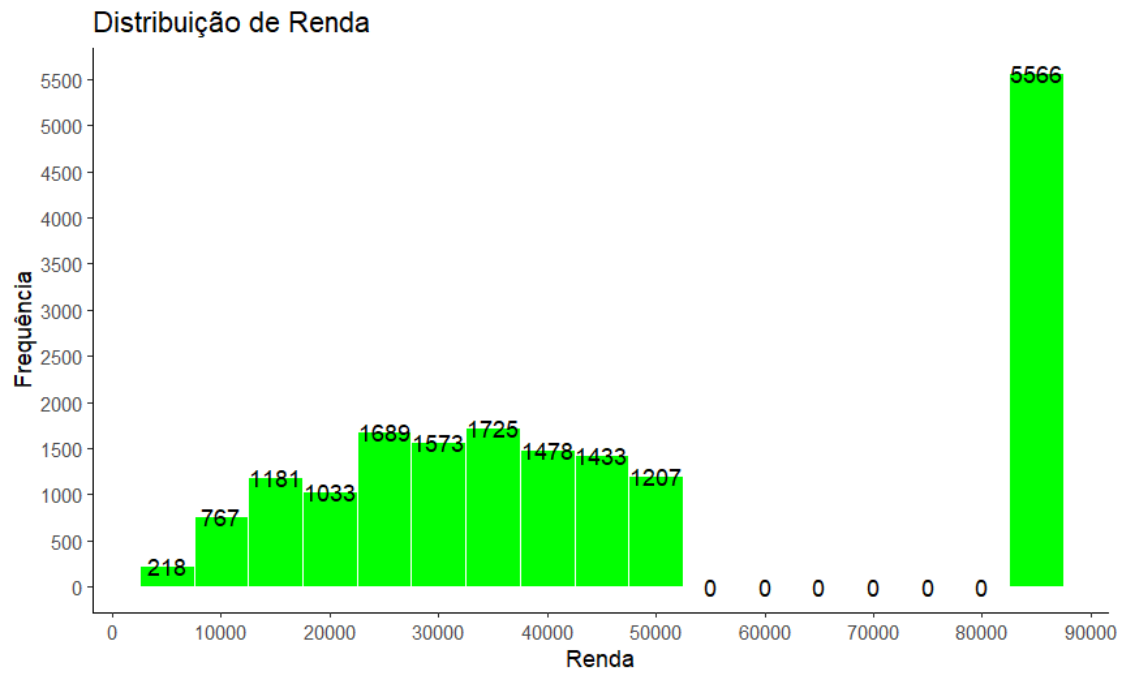
Sul: Delaware, Flórida, Geórgia, Maryland, Carolina do Norte, Carolina do Sul, Virgínia, Washington D.C., Virgínia Ocidental, Alabama, Kentucky, Mississippi, Tennessee, Arkansas, Louisiana, Oklahoma e Texas.

Oeste: Arizona, Colorado, Idaho, Montana, Nevada, Novo México, Utah, Wyoming, Alasca, Califórnia, Havaí, Oregon e Washington.

Ver o mapa disponível em www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf.



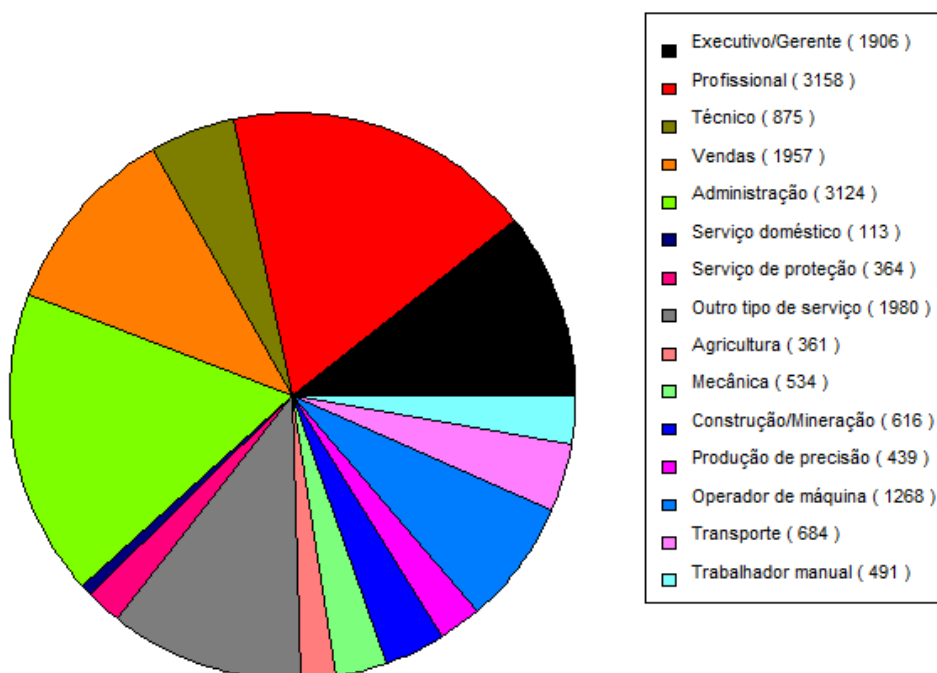




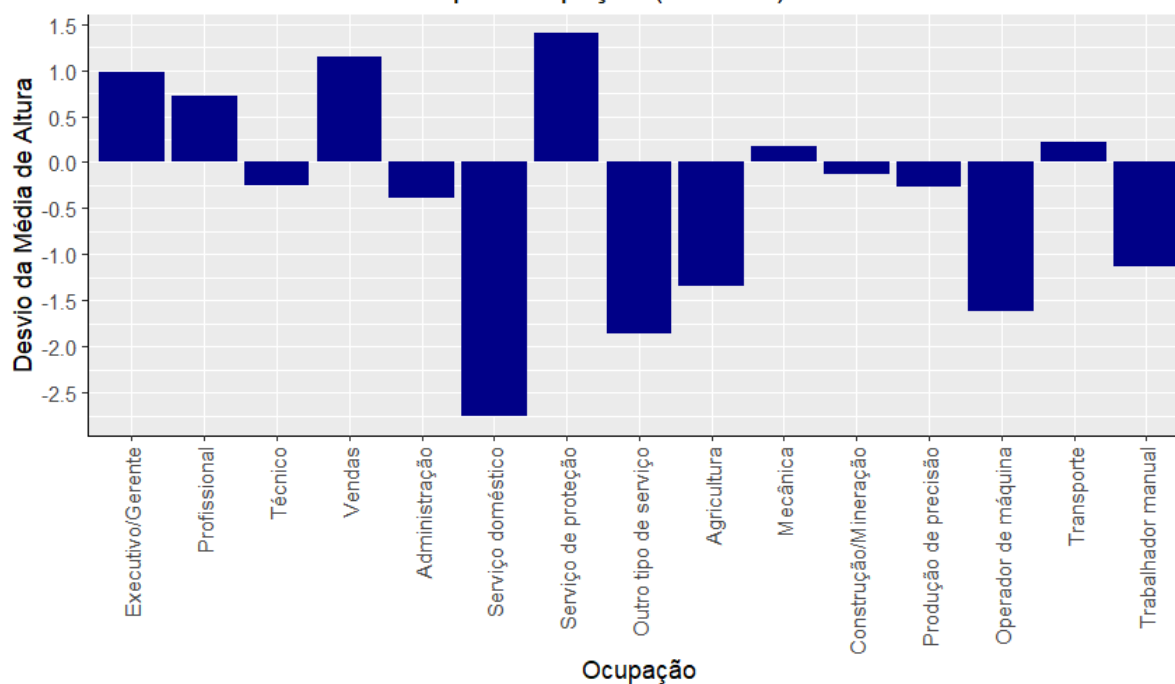
Dada a distribuição anormal das rendas na base, especialmente no que tange a última categoria de rendimentos (US\$ 84.054,75/ano), na seção de especificações, teremos regressões que não incluem esta categoria, dada a sensibilidade do OLS a *outliers*.

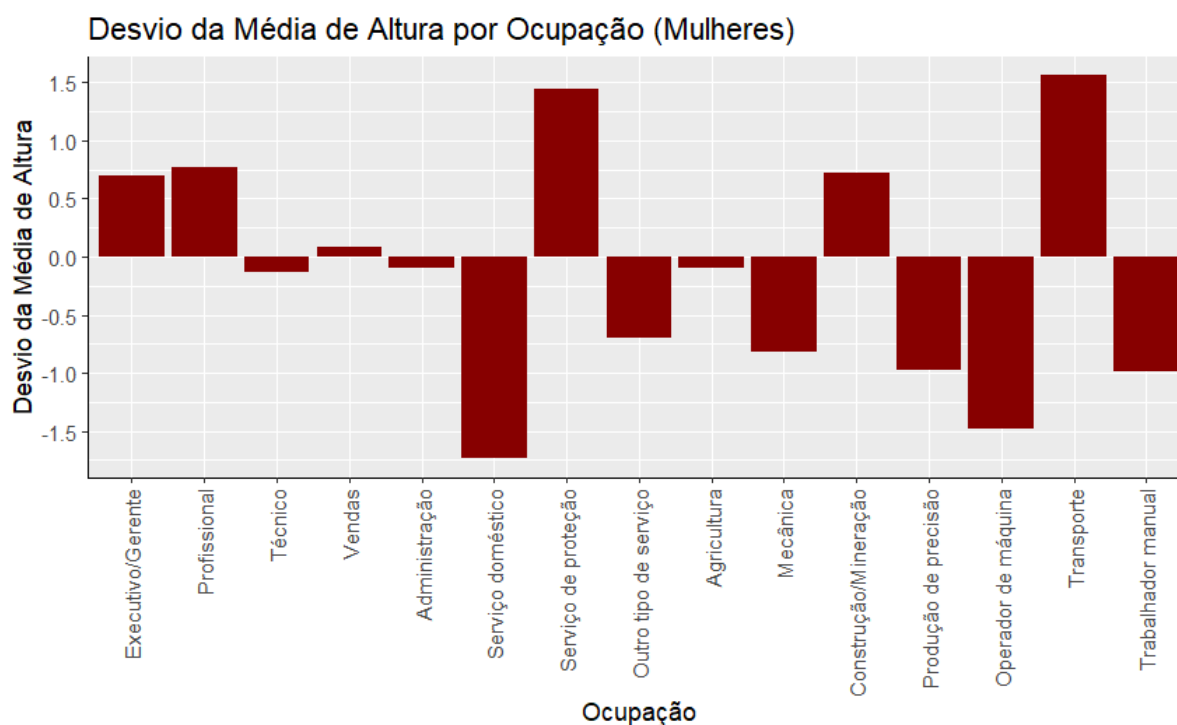
Outra variável interessante é ocupação:

Ocupação do Trabalhador

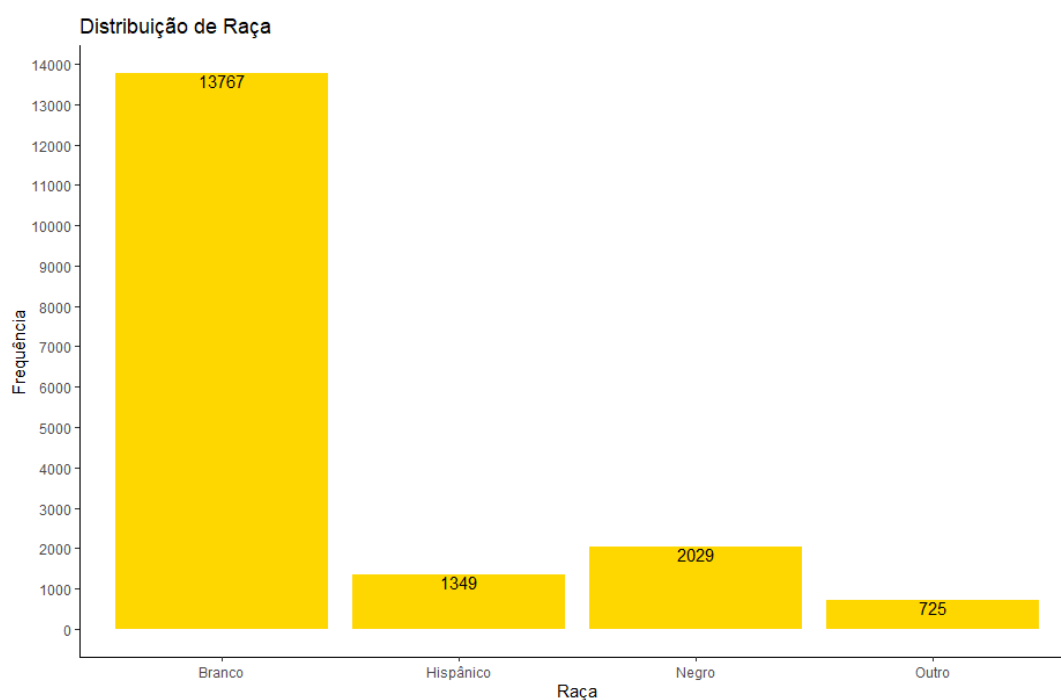


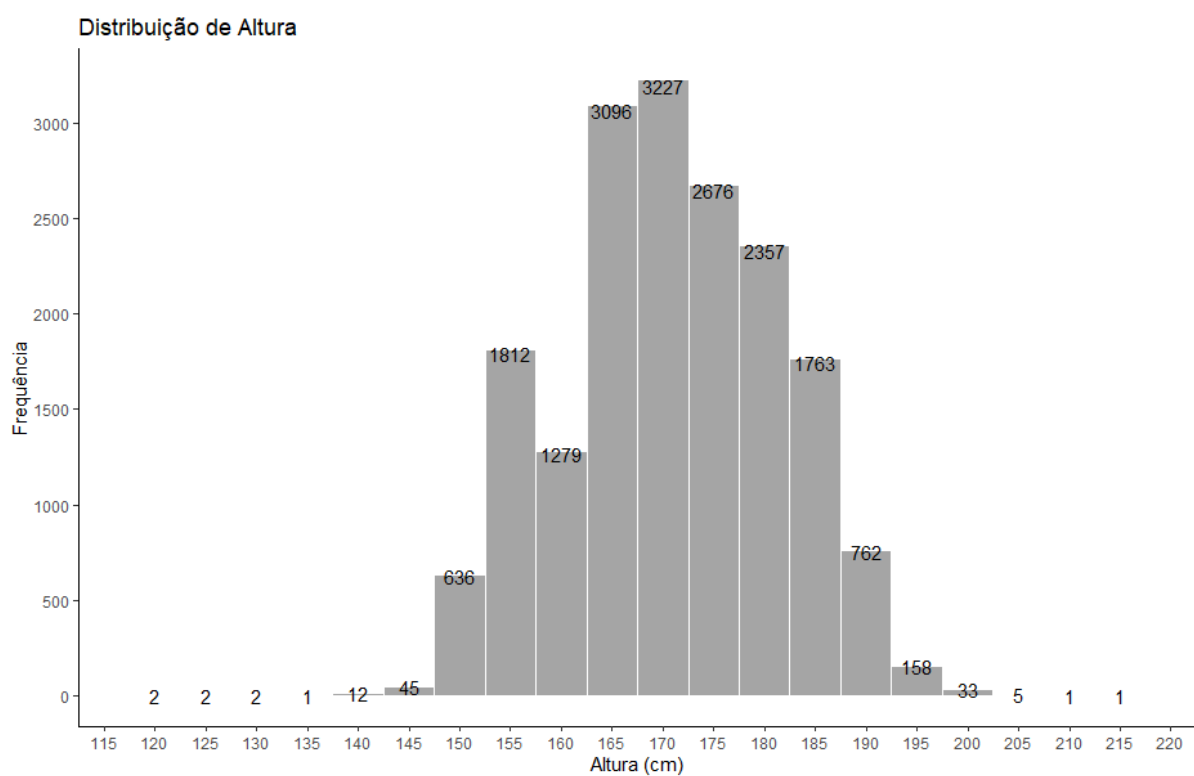
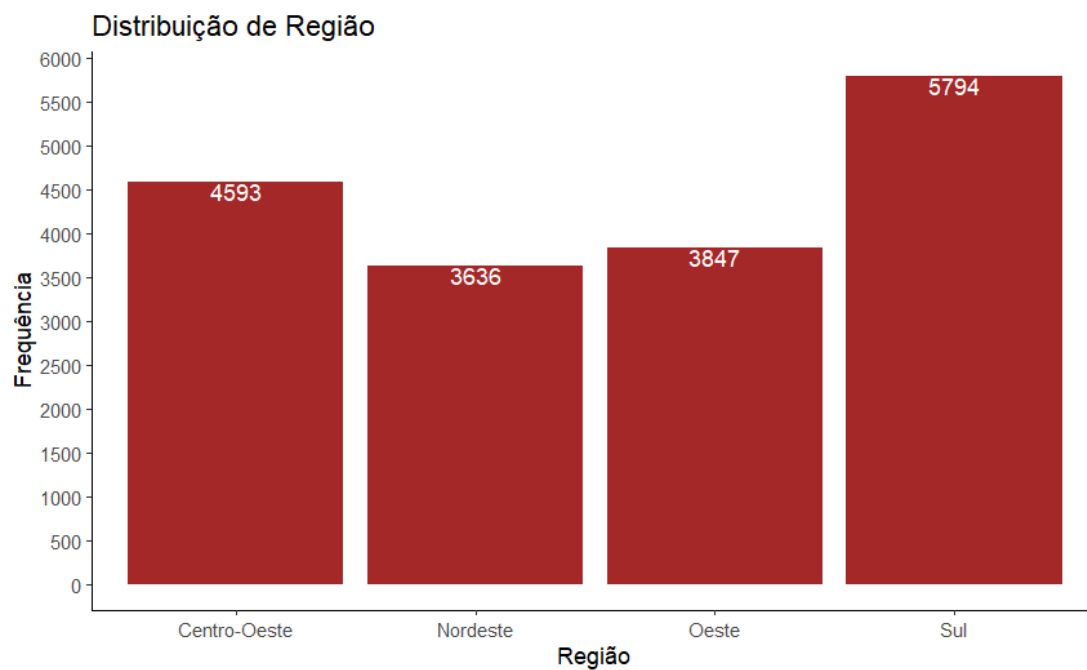
Desvio da Média de Altura por Ocupação (Homens)

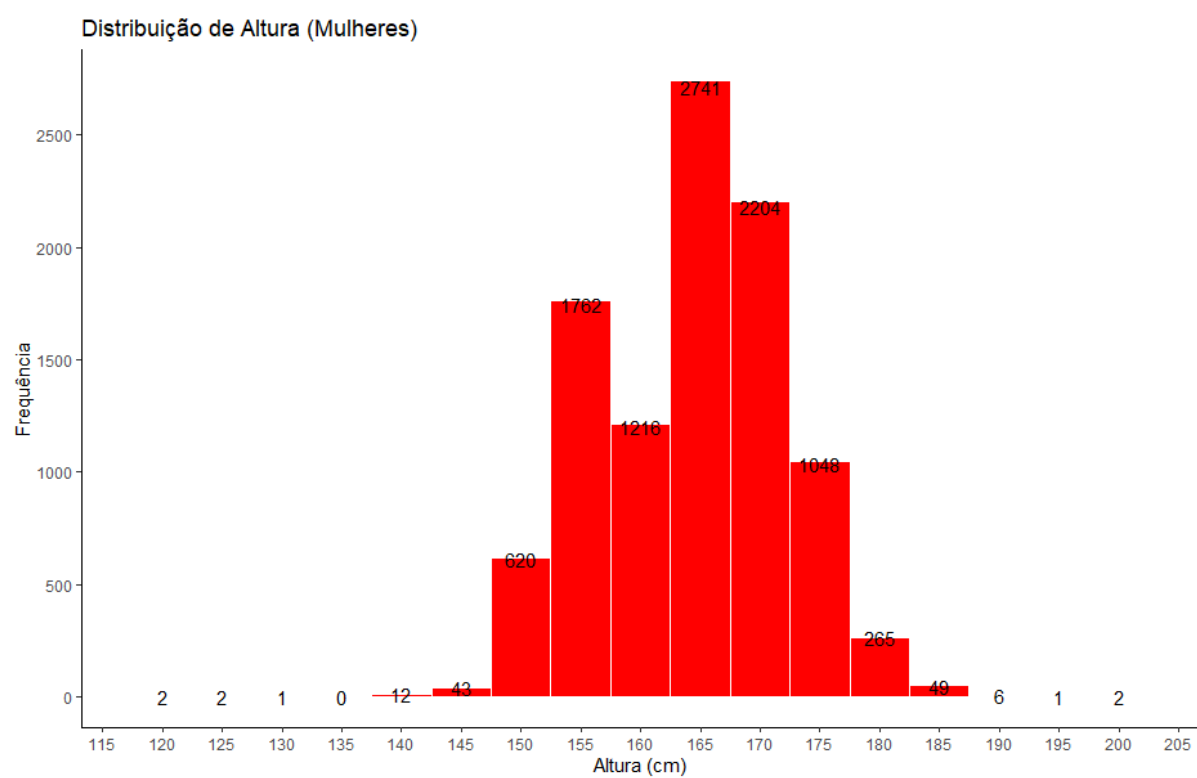
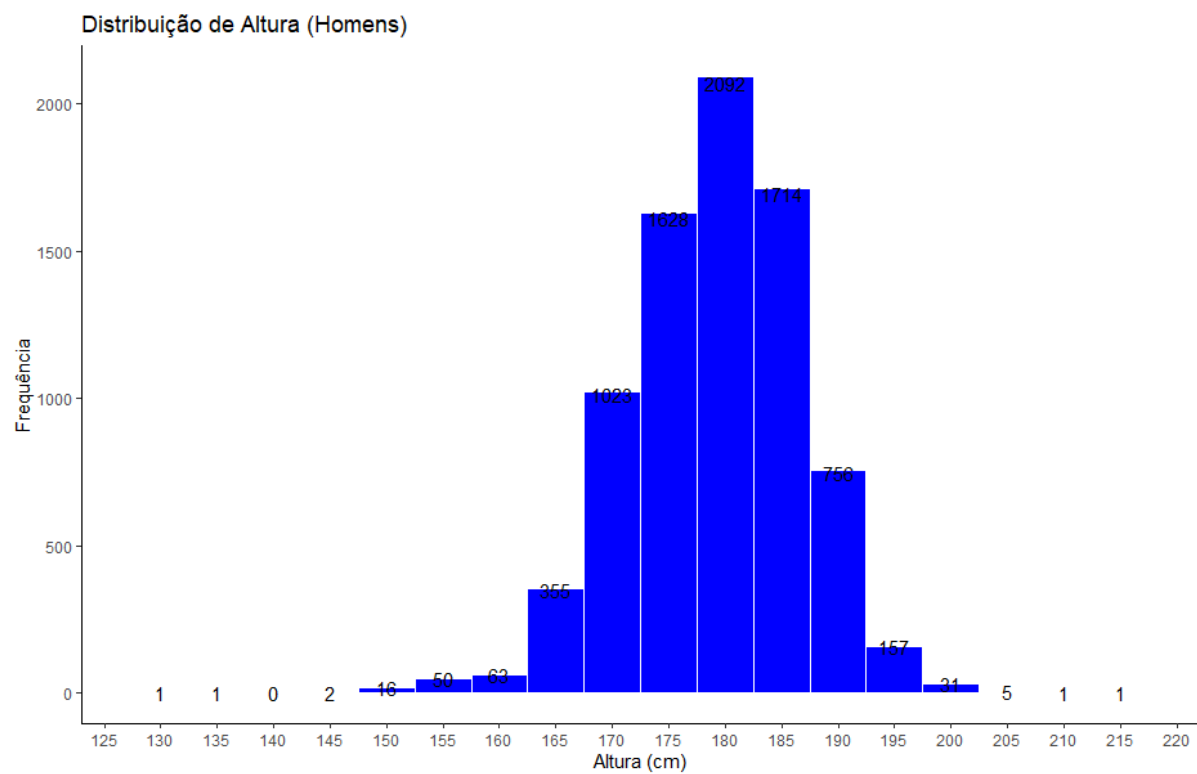


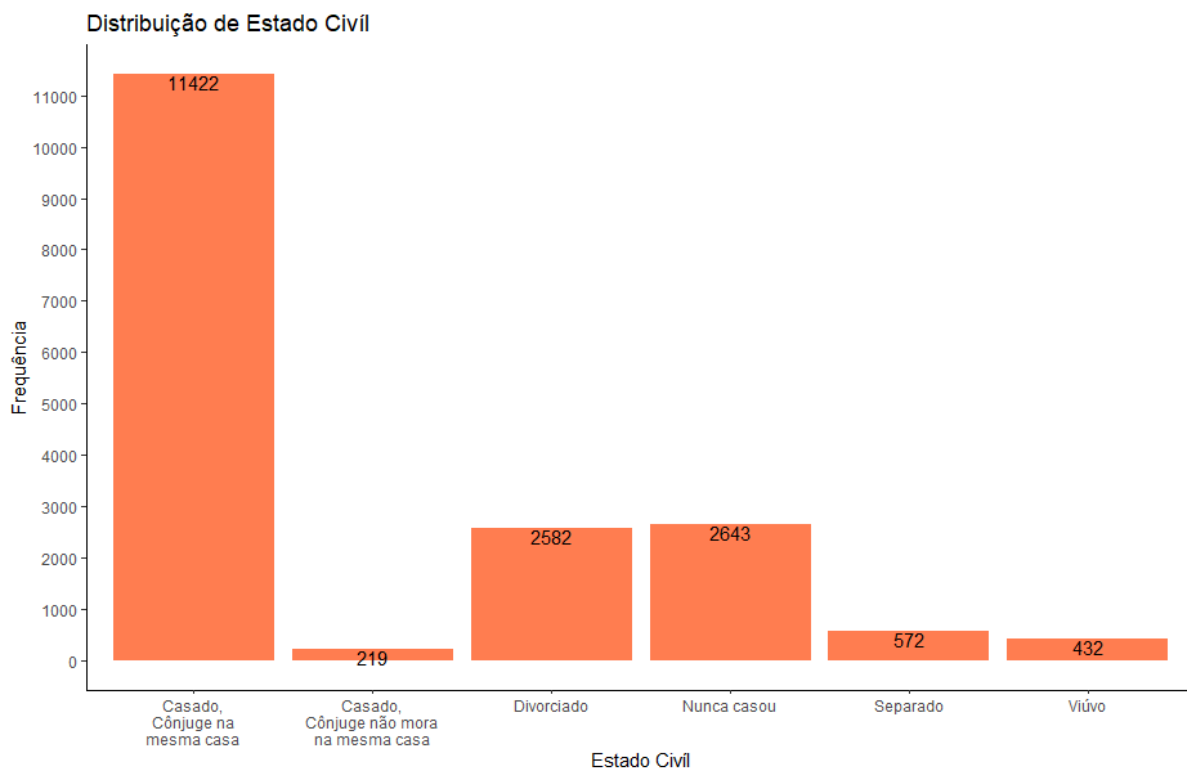
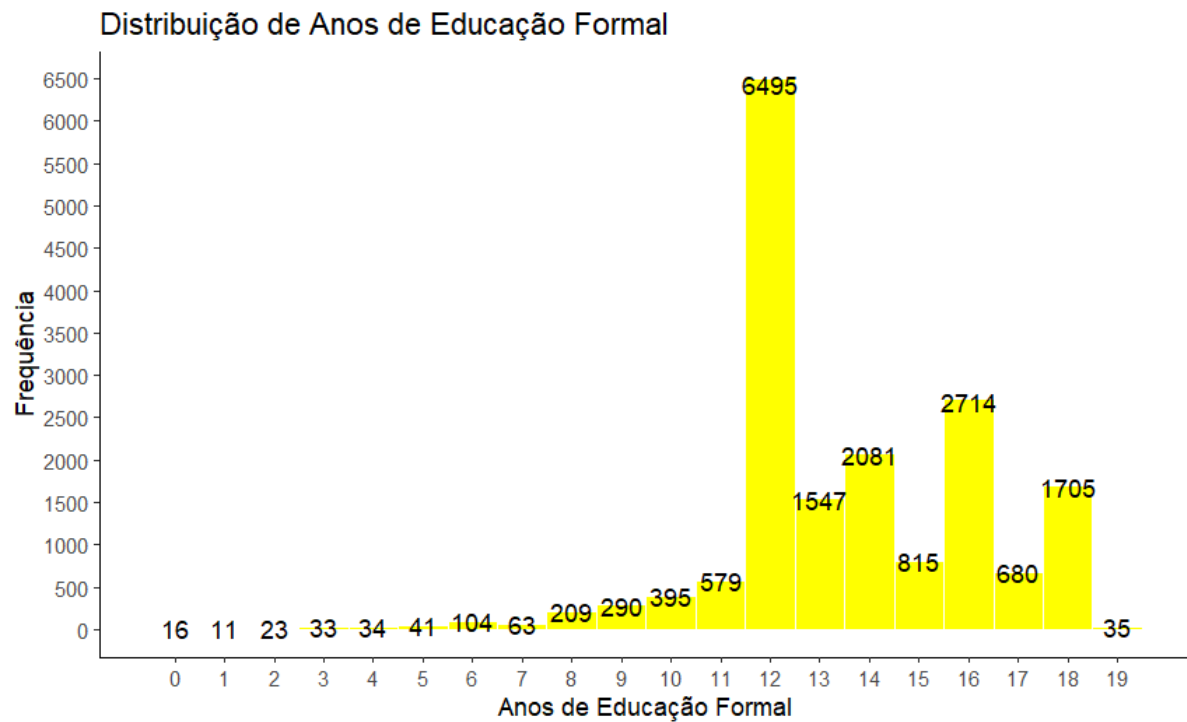


Os dois gráficos acima são baseados na figura 1 de Case e Paxson (2008). Interessa citar que nossos resultados são pertinentes para separar as ocupações manuais e intelectuais, em termos de desvio da altura média de cada sexo. Há exceções, como serviços de proteção, e transporte (pelo menos para as mulheres). Mais precisamente, essa divisão parece ser entre as ocupações administrativas-gerenciais-profissionais (as primeiras cinco) e as manuais-técnicas (as demais).





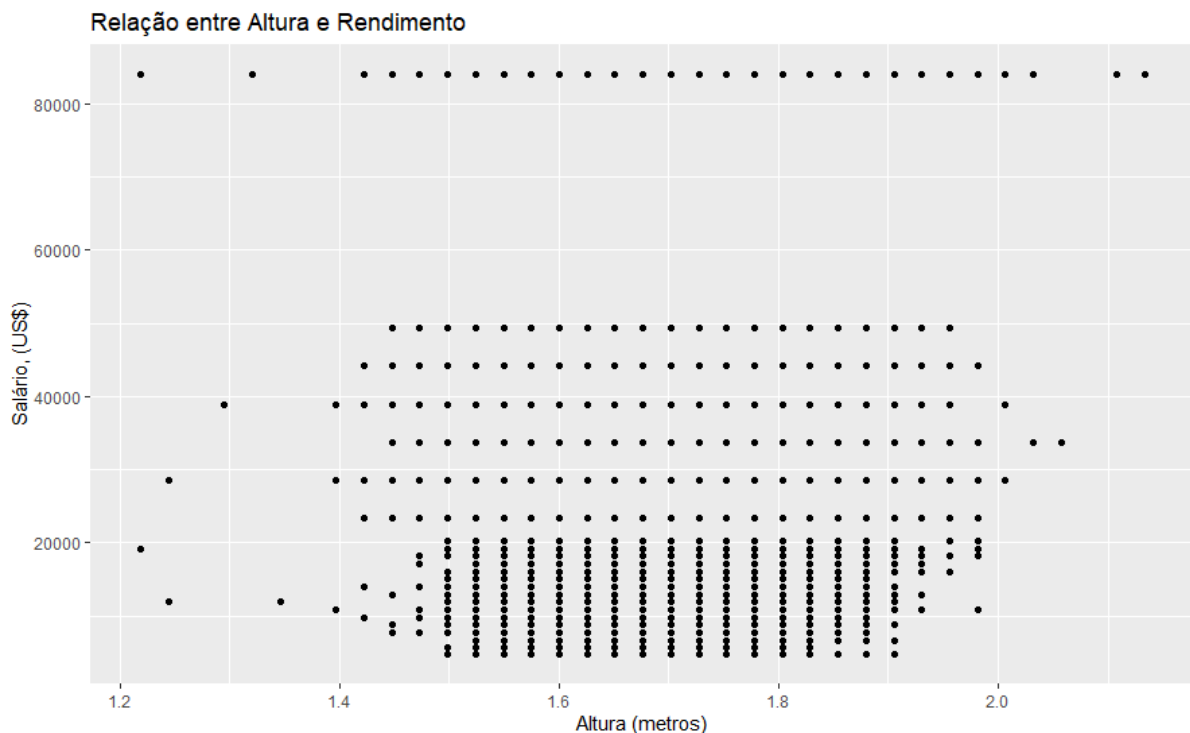




No caso das variáveis contínuas, usando a biblioteca *pander* elaboramos estatísticas descritivas:

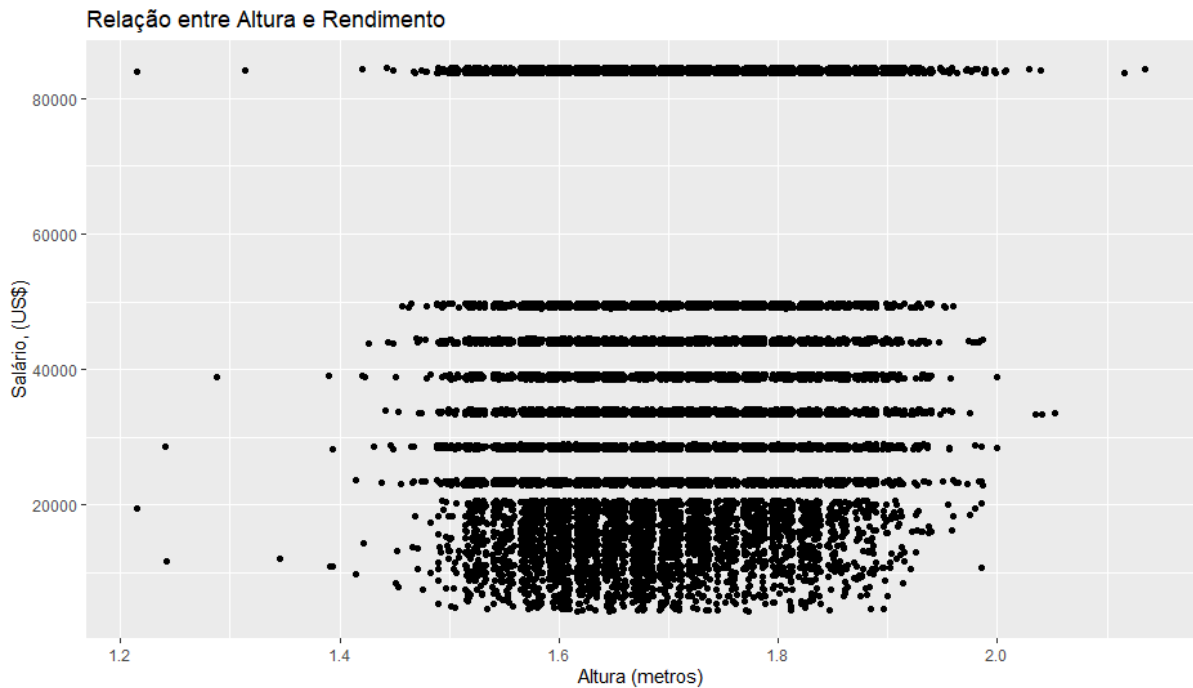
	age	educ	earnings	height_cm	weight_kg
Min.	25,00	0,00	4726,00	121,9	36,29
1st Qu.	33,00	12,00	23363,00	162,6	63,50
Median	40,00	13,00	38925,00	170,2	73,94
Mean	40,92	13,54	46875,00	170,1	77,27
3rd Qu.	48,00	16,00	84055,00	177,8	86,18
Max.	65,00	19,00	84055,00	213,4	227,25

Por fim, elaboramos uma nuvem de pontos para observar a relação entre altura e rendimentos³:



Note que, pela natureza discreta das variáveis dessa nuvem de pontos, ocorrem muitas sobreposições de pontos — “*overplotting*”. Isso dificulta a visualização e prejudica muito a intuição. Uma ferramenta disponibilizada pelo *ggplot2* é o *geom_jitter()*, que introduz uma pequena quantidade de variação aleatória na posição de cada um dos pontos, melhorando a visualização:

³ Para melhor entender este gráfico, especialmente no eixo dos rendimentos, é importante retomar o explicado na nota de rodapé 1.



3. Breve Revisão da Literatura

Começamos pelo artigo que nos forneceu a base de dados.

Case e Paxson (2008), inicialmente, indicam que, de fato, há correlação entre altura e salários. Encontram coeficientes de, mais ou menos, 0,015 para a altura, em relação ao logaritmo dos salários por hora. Em Português Brasileiro corrente, um aumento de uma polegada (2,54 cm) de altura é relacionado a um de 1,5% no salário. Entretanto, as autoras creem que esse “prêmio” por altura é devido — como mencionamos em 1.2. —, à altura ser meramente uma procuração para a habilidade cognitiva, que de fato é responsável pelos retornos salariais. A análise dessa premissa começa criando um modelo linear onde a habilidade cognitiva é uma função duma tal *dotação* (“endowment”) individual, que engloba os “efeitos combinados de condições ambientais (como saúde e nutrição), fatores biológicos, fatores genéticos e interações gene-ambiente.” (tradução nossa) (p. 508). Daí assumem que o salário é função linear da dotação, e que é o único fator que correlaciona habilidade cognitiva, altura e salários. Demonstram primeiro que, na média, crianças mais altas têm melhor desempenho em testes padronizados. Também mostram que trabalhadores mais altos são mais comuns nos setores onde capacidades cognitivas são mais relevantes. Analisando dados referentes a resultados de testes cognitivos, elas concluem que “os resultados dos testes são, em conjunto, altamente significativos [...], com testes F de 31,1 para homens e 38,5 para mulheres. A inclusão dessas pontuações em testes cognitivos reduz o tamanho dos

coeficientes de altura em mais de 50% e os torna estatisticamente insignificantes. Tanto os resultados dos testes como a altura estão correlacionados com os antecedentes familiares, e é interessante ver até que ponto os antecedentes familiares podem explicar os prêmios de altura.” e que “Esta redução nos prêmios de altura, na presença de controles alargados, não é surpreendente: se a altura é um marcador de capacidade cognitiva, então a inclusão de determinantes da capacidade cognitiva [...] deverá enfraquecer a associação entre altura e rendimentos.” (tradução nossa) (p. 521).

Também analisaremos dois estudos brasileiros. Curi e Menezes-Filho (2008) chegam a uma conclusão parecida, analisando dados brasileiros. Eles encontram que “a altura tem efeitos positivos e significantes na conclusão dos ciclos escolares.” e que “Em média, as pessoas mais altas têm nível educacional superior [...]. Os resultados do modelo mostram que as pessoas mais altas têm maior probabilidade de trabalhar em ocupações que exigem um nível de qualificação maior, como empregador, enquanto as pessoas mais baixas têm maior probabilidade de trabalhar em ocupações de pouca qualificação” (p. 19). Suliano, Irffi e Corrêa (2017) corroboram, com dados brasileiros específicos a mulheres, mas encontram que o “efeito da altura nos ganhos se dá até o ponto de máximo de 1,72 m [...]; após esse ponto, a altura tem um efeito negativo nos ganhos laborais.” (p. 217).

Hübler (2006) diverge, e chega a duas conclusões: “(1) O efeito altura varia com uma tendência não linear. Além disso, foram detectados movimentos cíclicos. Quanto maior for a taxa de crescimento desfasada do PIB, maior será o efeito da altura sobre os salários. Durante uma recessão, o prêmio de altura é menor ou desaparece. [...] (2) Para além da tendência não linear, a relação individual entre salários e altura também é não linear. O efeito máximo para os homens está evidentemente acima da média da altura masculina, enquanto o valor máximo análogo para as mulheres está abaixo da altura média feminina. A preferência por pessoas altas deve-se em parte à dotação e em parte à discriminação ou a efeitos de produtividade não observados que diferem entre trabalhadores baixos e altos. Para os homens, o efeito dotação é mais significativo, enquanto para as mulheres observamos o contrário. Em ambos os casos, a discriminação parece ter efeito. Encontramos discriminação por parte dos empregadores e menos obviamente discriminação por parte dos clientes. No geral, não existe uma ligação simples e univariada entre altura e salário, mas sim uma rede complexa de influências que foram detectadas.” (p. 18).

4. Estimações e Regressões

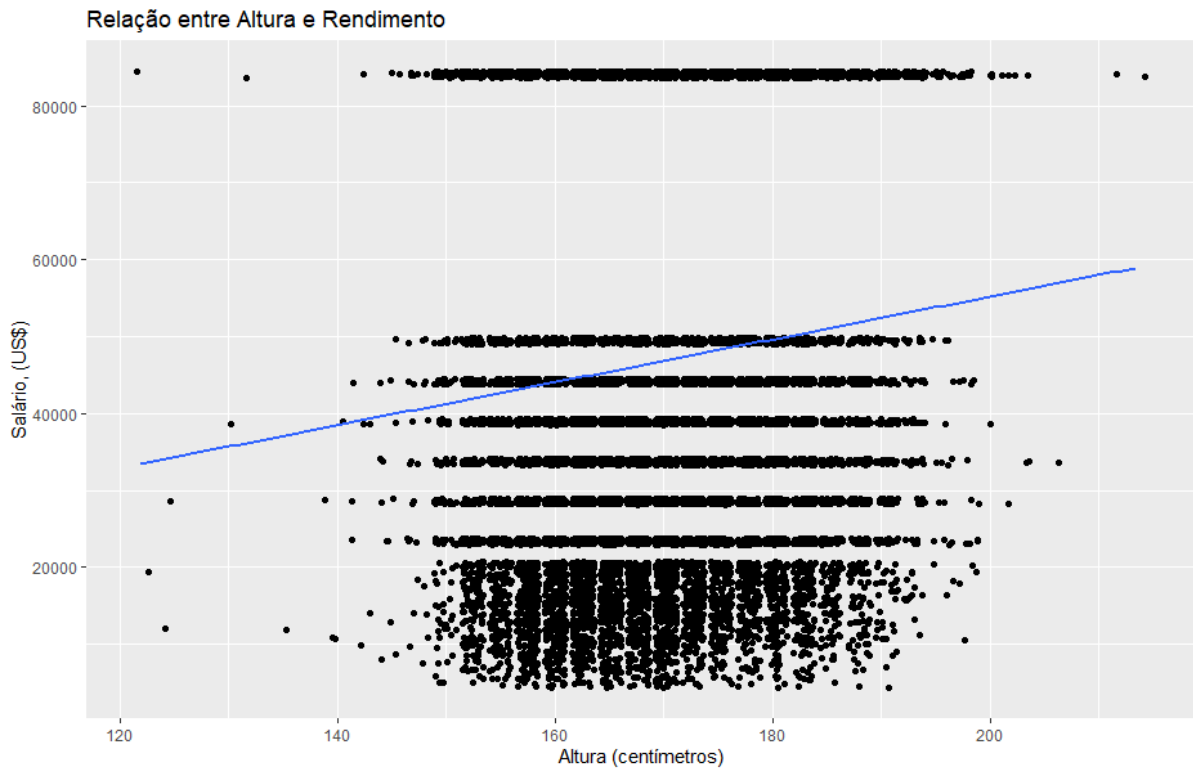
4.1. Regressão Preliminar

Regressão Preliminar	
	<i>Variável dependente:</i>
	Salário
Altura (cm)	278.611*** (19.878)
Constante	-512.734 (3,386.856)
Observações	17,870
R ²	0.011
R ² ajustado	0.011
Estatística F	196.456*** (df = 1; 17868)
<i>Nota:</i>	*p<0.1; **p<0.05; ***p<0.01
SE robusto a heterocedasticidade: (19.8406)	

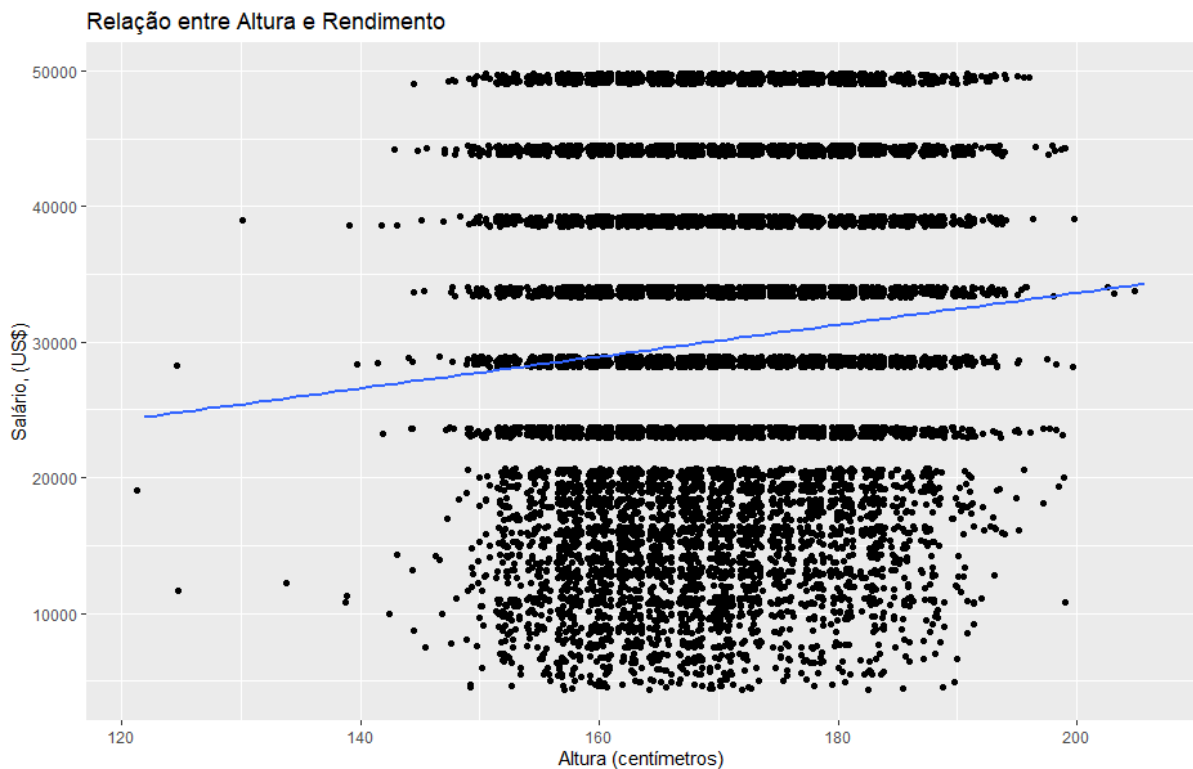
Os números de salários anuais resultam em um coeficiente homericamente grande — não obstante, o efeito marginal, significativo a 1%, parece razoável: um acréscimo de um centímetro de altura causaria um aumento de 278 dólares nos rendimentos anuais se respeitadas todas as hipóteses de inferência causal do OLS. Portanto, confirma-se a esperada relação positiva entre altura e rendimentos. É importante ressaltar que o intercepto não possui significância. Notamos, ainda, que o R^2 é baixo: apenas 1,1% da variação dos rendimentos anuais é explicada pela variação da altura. Isso é um forte indicativo de que há outras variáveis que determinam o salário de um indivíduo que não sua altura.

4.1.2. Variância dos Erros

Mesmo utilizando o recurso do `geom_jitter()` para melhorar a visualização, ainda sofremos com problemas de *overplotting* especialmente na porção superior da nuvem de pontos, o que torna um pouco dúbia a análise da variância dos erros:



Todavia, feita uma regressão eliminando os *outliers* em renda⁴, fica clara a heterocedasticidade dos erros:



⁴ Essa regressão será melhor explicada e desenvolvida na seção 4.2.2.

Note que nas menores alturas, os erros têm pouca variância. No entanto, a partir de 140cm de altura, a variância assume outro comportamento completamente diferente, anulando a plausibilidade de erros homocedásticos para esses dados. Isso implica primeiramente que as estimações de OLS não terão as propriedades desejáveis do Teorema de Gauss-Markov, e em segundo lugar que desvios-padrão robustos a heterocedasticidade deverão ser computados a fim de aferir significância.

4.1.3. Plausibilidade das Hipóteses de Causalidade do OLS

São três as hipóteses sob as quais o OLS é causal:

- I. $E[u/X] = 0$;
- II. (X_i, Y_i) são identicamente e independentemente distribuídos (i.i.d.);
- III. O modelo verdadeiro é linear e $\text{Var}(X) > 0$.

A primeira hipótese é a mais forte e é inválida para esta regressão. Inúmeros fatores determinam os rendimentos obtidos por um trabalhador, e isso causa correlação do erro com o regressor — isso será mais bem elaborado em breve.

A segunda hipótese, por outro lado, está relacionada com a amostragem dos dados — ela é verdadeira toda vez que os dados forem coletados via amostragem aleatória simples. Como citado anteriormente, a base de dados vem de uma pesquisa do NHIS e não foi feita a partir de uma amostra aleatória simples, e sim por amostragem via clusters geográficos. Isso traz risco de viés e coloca em jogo a segunda hipótese, sendo provável que ela de fato não seja atendida.

A terceira hipótese, no que tange à variância da altura, ela é, de fato, maior que zero. No entanto, linearidade nos parâmetros populacionais pode ser considerada uma hipótese forte para essa relação, como sugerido por estudos anteriores, vulgo Hübler (2006). Não obstante, esta é a hipótese mais provável de ser válida neste contexto. Na literatura, também encontramos como terceira hipótese ‘grandes *outliers* serem improváveis’: como mostrado na seção 2., a distribuição da renda na base de dados não é usual, isto é, não representativa da população estadunidense, com uma quantidade extremamente grande de observações na categoria de maior renda — para essa definição alternativa da hipótese III, confirmamos sua violação.

4.2. Regressões Lineares Simples com Restrições em Variáveis de Interesse

4.2.1. Gênero

A primeira variável de interesse que exploraremos será o gênero. A literatura possui evidências exaustivas em *gender pay gap* e ele também é observado na base de dados. Aplicando regressões separadas para homens e mulheres, temos:

Regressões por Sexo

	<i>Variável dependente:</i>	
	Salário	
	Homens	Mulheres
Altura (cm)	514.512*** (39.672)	201.269*** (38.936)
Constante	-43,130.340*** (7,068.481)	12,650.860** (6,383.741)
Observações	7,896	9,974
R ²	0.021	0.003
R ² ajustado	0.021	0.003
Estatística F	168.201*** (df = 1; 7894)	26.721*** (df = 1; 9972)
<i>Nota:</i>		*p<0.1; **p<0.05; ***p<0.01
SE robusto a heterocedasticidade:	(38.920)	(38.4191)
	Homens	Mulheres

Note que, ao separar os gêneros, o efeito marginal vai de 278,6 para toda a base, para 514,5 para homens, e 201,3 para mulheres. Ou seja, o efeito marginal da altura no rendimento anual é muito maior para homens do que para mulheres, conclusão já esperada.

4.2.2. Renda

A próxima variável a ser restrita é a própria renda anual. Por conta da sensibilidade do OLS a *outliers* e da distribuição não usual da renda na base, aplicamos uma regressão filtrando as observações em que a renda faz parte da maior classe:

Regressão sem outliers na renda

	<i>Variável dependente:</i>
	Salário
Altura (cm)	117.214*** (10.783)
Constante	10,183.520*** (1,831.374)
Observações	12,304
R ²	0.010
R ² ajustado	0.009
Estatística F	118.161*** (df = 1; 12302)
<i>Nota:</i>	*p<0.1; **p<0.05; ***p<0.01
SE robusto a heterocedasticidade: (10.7597)	

Percebe-se uma queda expressiva no valor do efeito marginal. A relação permanece positiva, mas com uma força consideravelmente menor — isso dá a entender que a presença dos *outliers* estava superestimando o efeito da altura na renda.

4.2.3. Região

Outra variável que potencialmente impacta a variável dependente, é a região dos EUA da observação. O Nordeste estadunidense é a região mais rica e com maior custo de vida dentre as outras, e isso certamente é um dos determinantes dos rendimentos anuais do indivíduo observado. Portanto, fazemos duas regressões, uma que contém somente observações do Nordeste, e outra que contém apenas observações das demais regiões:

Regressão restrita ao nordeste vs às demais regiões

	<i>Variável dependente:</i>	
	Salário	
	(Nordeste)	(Demais regiões)
Altura (cm)	316.585*** (45.336)	278.973*** (22.036)
Constante	-2,507.727 (7,693.116)	-1,708.035 (3,758.581)
Observações	3,636	14,234
R ²	0.013	0.011
R ² ajustado	0.013	0.011
Estatística F	48.763*** (df = 1; 3634)	160.266*** (df = 1; 14232)
<i>Nota:</i>		*p<0.1; **p<0.05; ***p<0.01
SE robusto a heterocedasticidade:	(44.6485) Nordeste	(22.0649) Demais regiões

Notamos uma mudança positiva expressiva no efeito marginal quando fixamos a região Nordeste, como esperado.

4.2.4. Anos de educação formal

Outra variável importante para explicar os rendimentos de um indivíduo, é a educação que ele recebeu. A literatura nessa relação é extensa e conhecemos uma relação majoritariamente positiva entre educação formal e rendimentos. Assim sendo, dividimos a base de dados entre aqueles que tiveram 10 ou menos anos de educação, e aqueles que tiveram mais de 10 anos de educação (maioria da base de dados):

Regressão restrita a mais de 10 anos de educação vs 10 ou menos

	<i>Variável Dependente:</i>	
	Salário	
	(>10 anos de educação)	(≤10 anos de educação)
Altura (cm)	238.128*** (20.612)	287.068*** (51.445)
Constante	7,837.569** (3,515.759)	-21,832.710** (8,633.528)
Observações	16,651	1,219
R ²	0.008	0.025
R ² ajustado	0.008	0.024
Estatística F	133.473*** (df = 1; 16649)	31.138*** (df = 1; 1217)
<i>Nota:</i>		*p<0.1; **p<0.05; ***p<0.01
SE robusto a heterocedasticidade:	(20.6006)	(56.8089)
	>10 anos	≤10 anos

Não obtivemos uma mudança tão expressiva como na restrição de outras variáveis, mas pudemos notar que o efeito da altura na renda é ligeiramente menor para aqueles com mais anos de educação, e ligeiramente maior para aqueles que possuem menos.

4.2.5. Classe

Também consideraremos a variável da classe do trabalhador. Especificamente, testamos se a relação altura-rendimentos difere para autônomos. Quiçá a ausência de RHs pode minimizar o impacto de discriminação.

Regressão Isolando Autônomos

	<i>Variável Dependente:</i>	
	Salário	
	Demais Classes	Autônomos
Altura (cm)	277.049*** (20.751)	324.631*** (70.414)
Constante	-113.903 (3,532.131)	-9,896.410 (12,128.080)
Observações	16,383	1,487
R ²	0.011	0.014
R ² Ajustado	0.011	0.013
Estatística-F	178.254*** (df = 1; 16381)	21.255*** (df = 1; 1485)
<i>Nota:</i>		*p<0.1; **p<0.05; ***p<0.01
SE Robusto a Heterocedasticidade:	(20.7275)	(69.2976)
	Demais Classes	Autônomos

Percebe-se haver uma diferença significativa entre os efeitos marginais e entre as constantes, que não possuem significância. Também interessa citar a proximidade dos R².

4.2.6. Raça

Outra questão interessante é a variável raça. Considerando as diferenças de condições socioeconômicas de base racial, consideramos inicialmente que o salário dos trabalhadores brancos deve ser maior. Portanto, os isolamos em comparação aos outros grupos raciais:

Regressão Brancos vs Não-Brancos

	<i>Variável Dependente:</i>	
	Salário	
	Não-Brancos	Brancos
Altura (cm)	230.095*** (40.655)	216.897*** (22.665)
Constante	-231.848 (6,825.244)	12,369.460*** (3,878.553)
Observações	4,103	13,767
R ²	0.008	0.007
R ² Ajustado	0.008	0.007
Estatística-F	32.032*** (df = 1; 4101)	91.578*** (df = 1; 13765)
<i>Nota:</i>		*p<0.1; **p<0.05; ***p<0.01
SE Robusto a Heterocedasticidade:	(41.8463)	(22.5918)
	Não-Brancos	Brancos

É evidente que o efeito marginal da altura difere muito pouco, mas é interessante notar a constante. O fato de a dos brancos ser positiva e significativa confirma o que foi indicado pelas médias salariais de cada raça. Também se nota que o R² não difere entre os grupos.

4.2.7. Estado Civil.

Examinando os dados, notamos que os trabalhadores casados possuíam média de rendimento maior que os não-matrimoniados. Assim, examinaremos essa diferença por regressões especializadas:

Casados vs Não-Casados

	<i>Variável Dependente:</i>	
	Salário	
	Não-Casados	Casados
Altura (cm)	339.929*** (27.434)	223.020*** (24.175)
Constante	-24,518.370*** (4,669.026)	16,619.280*** (4,121.762)
Observações	6,448	11,422
R ²	0.023	0.007
R ² Ajustado	0.023	0.007
Estatística-F	153.527*** (df = 1; 6446)	85.104*** (df = 1; 11420)
<i>Nota:</i>		*p<0.1; **p<0.05; ***p<0.01
SE Robusto a Heterocedasticidade:	(27.7591)	(24.0255)
	Não-Casados	Casados

Interessa notar que o efeito marginal da altura e o R² são significativamente maiores para os não casados. E que a constante é homericamente menor, confirmando o que sabemos sobre a média de rendimentos para esses trabalhadores.

4.2.8. Ocupação

Como visto nas tabelas de desvio da média de altura por ocupação, essa variável parece ser de alguma importância. Isso sem contar como é tratada em Hübler (2006) e Case e Paxson (2008). Por isso, usando a divisão estabelecida por nós, estimamos:

Adm.-Geren.-Prof. vs Manuais-Técnicas

	<i>Variável Dependente:</i>	
	Salário	
	Adm.-Geren.-Prof.	Manuais-Técnicas
Altura (cm)	326.687*** (25.847)	452.138*** (27.483)
Constante	-2,149.519 (4,379.079)	-40,740.750*** (4,724.952)
Observações	11,020	6,850
R ²	0.014	0.038
R ² Ajustado	0.014	0.038
Estatística-F	159.748*** (df = 1; 11018)	270.655*** (df = 1; 6848)
<i>Nota:</i>		*p<0.1; **p<0.05; ***p<0.01
SE Robusto a Heterocedasticidade:	(25.6354) Adm.-Geren.-Prof.	(27.889) Manuais-Técnicas

Evidente que a relação altura-rendimentos varia conforme a categoria de ocupação, vide os coeficientes. Mais interessante é a diferença homérica na constante, que quiçá reflete as distintas médias de renda, maior na primeira categoria. Também é notável o R² maior no caso da segunda categoria, possivelmente porque em tarefas manuais a altura — como reflexo de força física — é mais útil.

4.3. O Inevitável Viés de Variável Omitida

Não nos parece mistério o porquê de numa regressão como a nossa haver VVO. É evidente que os fatores que determinam os rendimentos dum trabalhador são extremamente variados, como apontam os estudos citados. Em especial, a tabela 4 de Hübler (2006), cuja variável dependente é log-salários, inclui experiência prévia, tempo na empresa, tamanho da firma, escolaridade dos pais, entre outras variáveis às quais não temos acesso. Evidentemente, o R² de quase 0,89 indica que as incluir é eficiente. Nosso R² reduzido talvez indique a presença de VVO.

De todo modo, ainda que a altura possa servir de procuração a outras competências — como inteligência ou força física —, a dotação que Case e Paxson (2008) teorizam não explica por total o salário dum trabalhador. Assim, podemos reiterar a presença de VVO na regressão preliminar, por haver variáveis que simultaneamente são correlacionadas com a altura e determinantes da renda.

4.4. Variáveis a Serem Inseridas no Modelo para Conter o VVO

Dados os resultados que obtivemos ao restringir a regressão simples para as variáveis de interesse, e levando em conta a definição do viés de variável omitida, decidimos que serão inseridas ao modelo as seguintes variáveis:

1. Gênero, por meio de uma variável binária que assume 1 se a observação for um homem, e 0 caso contrário;
2. NE, por meio de uma variável binária que assume 1 se a observação for da região Nordeste estadunidense, e 0 caso contrário;
3. Educação, que representa a quantidade de anos de educação formal;
4. Branco, por meio de uma variável binária que assume 1 se a observação ser de raça branca, e 0 caso contrário.
5. Manual-Técnico, por meio de uma variável binária que assume 1 se a observação tiver ocupação de natureza manual ou técnica, e 0 caso contrário.

4.5. Regressões Múltiplas

Para essa seção, seguiremos com a abordagem de estimar uma regressão com *outliers* em renda e outra sem, a fim de avaliarmos esses efeitos nas estimações do OLS.

4.5.1. Regressão Múltipla COM *outliers* em Renda

Regressão Múltipla

<i>Variável dependente:</i>	
Salário	
Altura (cm)	76.083*** (25.6342)
Gênero	3,920.305*** (513.6491)
Branco	6,789.832*** (433.5026)
Educ	2,940.446*** (78.5887)
ManualTecnico	−9,535.438*** (434.8116)
NE	3,885.366*** (454.5901)
Constante	−9,969.654** (4,157.1104)
Observações	17,870
R ²	0.192
R ² ajustado	0.192
Estatística F	708.621*** (df = 6; 17863)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01 SEs computados com robustez a heterocedasticidade	

Note que, como esperado, o coeficiente da altura perdeu muito da sua magnitude. Não obstante, manteve seu sinal e continua significativa a 1%. Note ainda que as variáveis adicionadas para o modelo possuem coeficientes extremamente maiores em módulo do que β_1 , reforçando que o efeito de ser mais alto na renda, se não sendo nulo, é irrelevante perto de outras variáveis determinantes dos rendimentos. Em outras palavras, apesar da significância estatística do coeficiente, isso é evidência de não haver significância econômica nele.

O R² ajustado (muito próximo do R² sem ajuste) aumentou vertiginosamente, e a proximidade entre o R² e o R² ajustado indica que as variáveis adicionadas mais que compensaram a penalização do R² ajustado — agora, aproximadamente 20% da variação dos rendimentos está sendo explicada pelo modelo.

A Estatística F apresenta grande significância, assumindo um valor de quase 709, ou seja, os coeficientes são conjuntamente significantes.

Regressão Múltipla (Sem Outliers de Renda)

<i>Variável Dependente:</i>	
	Salário
Altura (cm)	41.845*** (14.1817)
Gênero	1,588.188*** (287.4343)
Branco	4,164.000*** (237.6162)
Educ	864.861*** (43.6034)
ManualTecnico	-3,469.958*** (237.3438)
NE	1,613.295*** (257.5616)
Constante	9,321.474*** (2,302.3441)
Observações	12,304
R ²	0.120
R ² Ajustado	0.120
Estatística-F	279.667*** (df = 6; 12297)

Nota:

*p<0.1; **p<0.05; ***p<0.01

SEs computados com robustez a heterocedasticidade

Excluindo os *outliers* em renda, o coeficiente de altura cai mais ainda, mas mantém sua significância. Os coeficientes das outras variáveis têm sinais esperados. Interessa mencionar que novamente a constante mudou de sinal nessa etapa, ganhou significância, e houve redução no R², possivelmente indicando que os fatores selecionados são mais relevantes em níveis mais altos de renda. Conforme o parágrafo anterior, a altura tem relevância reduzida — quiçá nula, pelo menos em comparação. O comentário sobre a proximidade entre o R² e o R² ajustado, se repete.

5. Análise dos Resultados

5.1. Tratamentos Para a Validade Interna dos Resultados

5.1.1. Regressões não lineares

5.1.1.1. Log Renda

Usualmente na literatura, a renda é modelada, como dependente, logarítmica — a exemplo de Case e Paxson (2008). Assim, verificaremos se há algum aumento de encaixe, em relação à primeira regressão da sessão passada:

Regressão Múltipla (Log Renda)

<i>Variável Dependente:</i>	
Log Salário	
Altura (cm)	0.002*** (0.001)
Gênero	0.097*** (0.013)
Branco	0.209*** (0.011)
Educ	0.072*** (0.002)
ManualTecnico	-0.249*** (0.011)
NE	0.101*** (0.011)
Constante	9.038*** (0.102)
Observações	17,870
R ²	0.210
R ² Ajustado	0.210
Estatística-F	790.494*** (df = 6; 17863)

Nota: *p<0.1; **p<0.05; ***p<0.01

SEs computados com robustez a heterocedasticidade

A regressão log-linear confirma e solidifica alguns resultados da primeira regressão múltipla. Em especial, o coeficiente de altura pequeno. Também notamos

um aumento no R² e na estatística-F. Há ainda o sinal mudado da constante. De todo modo, os resultados não diferem do esperado.

5.1.1.2. Altura quadrática para mulheres

A fim de testar o encontrado por Suliano, Irffi e Corrêa (2017), aplicamos um polinômio quadrático na altura ao restringir a regressão apenas para mulheres, e comparamos com a regressão linear correspondente:

Regressão múltipla restrita a mulheres

	<i>Variável Dependente:</i>	
	Salário	
	(Linear)	(Quadrado na altura)
Altura (cm)	−34.794 (35.7075)	
Altura ¹ (cm)		−680.138 (1236.41640)
Altura ² (cm)		1.966 (3.76025)
Branco	7,371.465*** (565.5963)	7,394.049*** (566.39052)
educ	3,015.264*** (107.3907)	3,017.731*** (107.33081)
ManualTécnico	−11,023.810*** (569.6840)	−11,029.450*** (569.62731)
NE	4,022.810*** (596.2766)	4,023.473*** (596.29713)
Constant	7,120.279 (5877.9146)	59,943.020 (101495.31238)
Observações	9,974	9,974
R ²	0.189	0.189
R ² ajustado	0.188	0.188
Estatística F	464.274*** (df = 5; 9968)	386.931*** (df = 6; 9967)

Nota:

*p<0.1; **p<0.05; ***p<0.01
SEs computados com robustez a heterocedasticidade

Percebemos de imediato que, restringindo a base para conter apenas mulheres, na regressão múltipla linear, o coeficiente da altura perde completamente sua significância, o R² cai e a estatística F se manteve significativa. Na regressão não linear, nada mudou em termos de significância, e o R² não apresentou melhora.

Concluímos, então, que não pudemos identificar os retornos à altura para mulheres de modo semelhante a Suliano, Irffi e Corrêa (2017).

5.1.2. Possíveis Erros de Medição nas Variáveis da Base

Acreditamos que os possíveis erros de medição nas variáveis da base podem estar relacionados a possíveis erros de aferição de medidas (como altura e peso) e possíveis erros de lembrança na definição de faixas de salário na pesquisa do NHIS. Entretanto, não há qualquer erro imediatamente óbvio em qualquer uma das variáveis da base.

5.1.3. Dados Faltantes

Não houve problemas relacionados a dados faltantes nessa base. Como ela é um subconjunto da base da pesquisa do NHIS, provavelmente os responsáveis por elaborar esse subconjunto já lidaram previamente com esses tipos de problemas.

5.2. Comentários Sobre Causalidade

De fato, não é possível afirmar que há relação causal da altura no salário, isto é, um aumento na altura gerar um aumento nos rendimentos. Como discutido por Case e Paxson (2008) e Hübler (2006), a altura pode bem ser um *proxy* para intelecto, havendo correlação entre altura e anos de educação, dentre outras variáveis mais relevantes na determinação do salário de um indivíduo. Nossos resultados demonstraram grande fragilidade da plausibilidade de efeito causal de altura na renda, tendo outros fatores como determinantes muito mais significativos. Como desenvolvido por Hübler (2006), ainda pode haver uma espécie de causalidade reversa no que tange a essas variáveis: as evidências apontam que pessoas mais altas assim o são por terem tido um melhor desenvolvimento na infância, por melhor qualidade na alimentação, dentre outros quesitos inerentes de uma renda familiar maior. E nesses casos, pessoas com mais altura, apenas emulariam o nível de renda familiar a qual foram previamente expostos.

5.3. Comentários Sobre Validade Externa

No que diz respeito à direção das relações encontradas entre as variáveis com significância econômica na determinação da renda, não é um grande salto afirmar que elas se mantêm em contextos diferentes dos EUA. Não obstante, as magnitudes dos coeficientes certamente não se aplicam a contextos mais

abrangentes, sendo eles intimamente relacionados com a estrutura específica da sociedade e do mercado de trabalho estadunidense.

6. Conclusão

Este estudo examinou a relação entre altura e salários, utilizando dados do US National Health Interview Survey (NHIS) de 1994. A análise foi baseada nos trabalhos de Case e Paxson (2008) e outros estudos correlatos, investigando se a altura tem algum efeito causal nos rendimentos de um indivíduo. Foram analisadas em conjunto com essas duas variáveis outras como gênero, anos de educação formal, raça, estado civil, região dos EUA e tipo de ocupação. Foram utilizadas regressões lineares, não lineares, univariadas e multivariadas para o estudo dessa relação.

Apesar da significância estatística encontrada no efeito marginal da altura nos rendimentos, principalmente nas regressões mais preliminares, não há plausibilidade de se cumprir as hipóteses de inferência causal do método Mínimos Quadrados Ordinários para esses dados. Além disso, nas regressões múltiplas, o efeito marginal da altura dificilmente tem alguma significância econômica.

Portanto, com base nas evidências encontradas, corroboramos estudos anteriores no que diz respeito a altura funcionar como uma procuração para características relevantes para a produtividade que por sua vez de fato impactam os rendimentos, mas não como um fator determinante em si só.

7. Referências Bibliográficas

Case, Anne; e Paxson, Christina (2008). *Stature and Status: Height, Ability, and Labor Market Outcomes*. Journal of Political Economy, 2008, 116(3): 499–532.

CENTERS FOR DISEASE CONTROL AND PREVENTION. *About the National Health Interview Survey*. Disponível em: https://www.cdc.gov/nchs/nhis/about_nhis.htm. Acesso em: 5 jun. 2024.

Curi, Andréa Zaitune; e Menezes-Filho, Naércio Aquino (2008). *A Relação Entre Altura, Escolaridade, Ocupação e Salários no Brasil*. Pesquisa e Planejamento Econômico, 38, no. 3.

Hübler, Olaf (2006). *The Nonlinear Link between Height and Wages: An Empirical Investigation*. IZA DP No. 2394.

Hlavac, Marek (2022). *Stargazer: Well-Formatted Regression and Summary Statistics Tables*. Pacote R versão 5.2.3. Disponível em: <https://CRAN.R-project.org/package=stargazer>.

Suliano, Daniel; Irffi, Guilherme; e Corrêa, Márcio Veras (2017). *Medidas Antropométricas e Retornos Salariais com Ênfase no Padrão de Atratividade e Saúde Corporal*. Pesquisa e Planejamento Econômico, 47, no. 2.