



Universidade de Brasília
Departamento de Estatística

Análise da diferença salarial no Distrito Federal

Talia Alves Xavier

Projeto apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

Brasília
2021

Talia Alves Xavier

Análise da diferença salarial no Distrito Federal

Orientador(a): Prof(a). Thais Carvalho Valadares Rodrigues

Projeto apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília
2021**

Agradecimentos

Primeiramente, agradeço à minha orientadora Thais Rodrigues por sua instrução, dedicação e paciência durante todo o projeto. Também agradeço a minha família e amigos por todo apoio e incentivo aos estudos, em especial à minha mãe, Marina Pereira, sem a qual meu progresso não seria possível, e ao Luiz Filipe da Silva. E, por fim, agradeço à minha amiga Larissa Gomes, pela companhia e apoio durante toda a jornada de graduação.

Resumo

Este trabalho estuda a diferença salarial entre indivíduos residentes no Distrito Federal no ano de 2018, levando em consideração características como sexo, raça, nível de escolaridade, local de moradia ou setor de trabalho. Para isso foi empregada uma regressão quantílica dos salários em função dessas características. Os resultados indicam que: mulheres recebem menos que os homens e essa diferença chega a 18,6% para o quantil 0,75 de salários; negros recebem menos que não negros em todos os quantis; pessoas com ensino superior completo têm maiores recompensas, principalmente, para os quantis elevados de salário; quem trabalha no setor público recebe salário maior em relação a quem não trabalhava no setor público, de forma que o aumento é mais percebido no menor nível de renda; quem trabalha na área de serviços domésticos apresenta menor recompensa do que quase todas as outras áreas em todos os quantis de salário e trabalhadores informais ganham menos se comparados aos formais nos quantis inferiores de renda, mas ganhavam mais no maior quantil de salários.

Palavras-chaves: Salário; Sexo; Regressão Quantílica; PDAD 2018.

Lista de Quadros

Lista de Tabelas

1	Taxas de desocupação segundo sexo, Distrito Federal - 2018	22
2	Medidas descritivas de idade e tempo no emprego, Distrito Federal - 2018 .	22
3	Portadores de deficiência segundo sexo, Distrito Federal - 2018	23
4	Posição no trabalho segundo sexo, Distrito Federal - 2018	26
5	Área de trabalho segundo sexo, Distrito Federal - 2018	26
6	Medidas descritivas do salário-hora (em reais), Distrito Federal - 2018 . . .	27
7	Regiões administrativas segundo nível de renda, Distrito Federal - 2018 . .	34
8	Resultados dos testes de Correlação de Kendall, Distrito Federal - 2018 . .	37
9	Resultados da regressão de trabalho, Distrito Federal - 2018	39
10	Pessoas que trabalham e não têm renda, segundo função no trabalho, Distrito Federal - 2018	40
11	Critérios de Akaike, Bayesiano e R^2 para seleção de variáveis segundo modelo usado, Distrito Federal - 2018	41
12	ANOVA para seleção de modelos, Distrito Federal - 2018	42
13	Resultados da regressão linear de $\log(\text{salários-hora})$, Distrito Federal - 2018	42
14	Resultados da regressão quantílica de $\log(\text{salários-hora})$, Distrito Federal - 2018	45

Lista de Figuras

1	Proporção de homens e mulheres, Distrito Federal - 2018	20
2	Pirâmide etária, Distrito Federal - 2018	21
3	Pessoas em idade de trabalhar x Pessoas que trabalham, segundo sexo, Distrito Federal - 2018	21
4	Raça segundo sexo, Distrito Federal - 2018	23
5	Portadores de deficiência grave segundo sexo, Distrito Federal - 2018	23
6	Residir com filhos segundo sexo, Distrito Federal - 2018	24
7	Nível de escolaridade segundo sexo, Distrito Federal - 2018	24
8	Estado civil segundo, Distrito Federal - 2018	25
9	Distribuição de salários-hora, Distrito Federal - 2018	27
10	Distribuição de salários-hora, Distrito Federal - 2018	28
11	Distribuição de salários-hora, segundo sexo, Distrito Federal - 2018	28
12	Distribuição de salários-hora, segundo raça, Distrito Federal - 2018	29
13	Distribuição de salários-hora, segundo porte de deficiência, Distrito Federal - 2018	29
14	Distribuição de salários-hora, segundo residência com filhos, Distrito Fede- ral - 2018	30
15	Distribuição de salários-hora, segundo nível de escolaridade, Distrito Fede- ral - 2018	30
16	Distribuição de salários-hora, segundo estado civil, Distrito Federal - 2018 .	31
17	Distribuição de salários-hora, segundo quantidade de trabalhos, Distrito Federal - 2018	31
18	Distribuição de salários-hora, segundo função no trabalho, Distrito Federal - 2018	32
19	Distribuição de salários-hora, segundo função no trabalho, Distrito Federal - 2018	32
20	Distribuição de salários-hora, segundo setor de trabalho, Distrito Federal - 2018	33

21	Distribuição de salários-hora, segundo setor de trabalho, Distrito Federal - 2018	34
22	Distribuição de salários-hora, segundo RA de moradia, Distrito Federal - 2018 (A)	35
23	Distribuição de salários-hora, segundo RA de moradia, Distrito Federal - 2018 (B)	35
24	Distribuição de salários-hora, segundo RA de moradia, Distrito Federal - 2018 (C)	36
25	Distribuição de salários-hora, segundo idade, Distrito Federal - 2018	36
26	Distribuição de salários-hora, segundo tempo no emprego, Distrito Federal - 2018	37
27	Matriz de Correlação de Kendall, Distrito Federal - 2018	37
28	Comparação entre quantis obtidos para os resíduos e os de uma Distribuição Normal, Distrito Federal - 2018	44
29	Coefficiente da regressão quantílica de $\log(\text{salário-hora})$ para sexo, Distrito Federal - 2018	48
30	Coefficiente da regressão quantílica de $\log(\text{salário-hora})$ para raça, Distrito Federal - 2018	48
31	Coefficiente da regressão quantílica de $\log(\text{salário-hora})$ para ensino superior, Distrito Federal - 2018	48
32	Coefficiente da regressão quantílica de $\log(\text{salário-hora})$ para setor informal, Distrito Federal - 2018	48
33	Estimativas e intervalos de confiança da regressão quantílica e da linear para $\text{Log}(\text{salário-hora})$, Distrito Federal - 2018 (parte 1)	55
34	Estimativas e intervalos de confiança da regressão quantílica e da linear para $\text{Log}(\text{salário-hora})$, Distrito Federal - 2018 (parte final)	56

Sumário

1 Introdução	8
2 Objetivos	10
2.1 Objetivo Geral	10
2.2 Objetivos Específicos	10
3 Revisão bibliográfica	11
3.1 Regressão Linear Múltipla	11
3.1.1 Razão Inversa de Mills	11
3.2 Regressão Quantílica	12
3.2.1 Quantis	12
3.2.2 Estimadores	14
3.2.3 Intervalos de Confiança	15
4 Materiais e métodos	17
4.1 Materiais	17
4.1.1 Variáveis	17
4.2 Métodos	18
5 Resultados	20
5.1 Análise Exploratória	20
5.1.1 Características dos indivíduos	20
5.1.2 Salários	27
5.2 Análise de Regressão Linear	38
5.2.1 Regressão de trabalho - Probit	38
5.2.2 Seleção de variáveis	40
5.3 Análise de Regressão Quantílica	44
6 Conclusão	49
Referências	51
Apêndice	52

A Seleção de variáveis - Regressão linear	52
B Gráficos: Regressão quantílica x Regressão linear	54

1 Introdução

A remuneração salarial é uma temática interessante de ser analisada, uma vez que está diretamente atrelada às questões de desigualdade social, pois um indivíduo que possui uma remuneração bem maior que a de um outro tende a ter melhor qualidade de vida.

A estruturação inicial da sociedade tem reflexos no seu desenvolvimento e nos papéis desempenhados pelos indivíduos que a compõem. Inicialmente pessoas do sexo feminino não possuíam direito de trabalhar, o que pode tê-las conduzido a um cenário de menor espaço no mercado de trabalho. Esse espaço também está relacionado à segregação ocupacional, que se dá pela distinção de trabalhos para homens e mulheres, e até 2014, segundo Sousa, Guedes e Rocha (2016), ainda era uma realidade desfavorável para elas no Brasil. Esse contexto histórico pode trazer consequências na recompensa pelo trabalho desempenhado.

Além disso, quando se pensa nas causas para essa diferença salarial, alguns aspectos são imediatos, como o nível de escolaridade, o local de trabalho e de moradia, o tipo de função e o setor no qual trabalho é desempenhado. Esses aspectos, em geral, podem ser medidos e podem explicar alguma parte dessa diferença salarial. Contudo, também há indícios de que outras características pessoais e não relacionadas ao contexto laboral possam influenciar essa distinção, como é o caso do sexo e raça/cor.

Quanto a isso, Ribeiro, Komatsu e Filho (2020) analisaram esses diferenciais no Brasil e verificaram que, para o período de 2016 a 2018, o tipo de instituição de ensino, gênero e raça eram significantes, de modo que pessoas que cursaram ensino médio em escola privada e superior em pública, são do sexo masculino e brancas apresentam melhores salários (correspondendo a 114% a mais de salários de mulheres não brancas com ensino médio público e superior privado). Além do mais, a recompensa de estudar em cada tipo de ensino é percebida de forma diferente entre os grupos de gênero e raça. Em todas as combinações de ensino, a maior remuneração era a de homens brancos, em seguida, estavam os homens não brancos, mulheres brancas e, por último, as mulheres não brancas (RIBEIRO; KOMATSU; FILHO, 2020).

Tendo isso em mente, este trabalho objetiva verificar de que forma a diferença salarial entre homens e mulheres da área urbana do Distrito Federal ocorre, identificando possíveis causas para isso, e qual é a sua magnitude. Para isso, será implementada uma regressão quantílica dos salários em função de características pessoais como sexo, raça/cor,

escolaridade e presença de filhos. O método de análise escolhido permite uma análise mais ampla (KOENKER, 2005), porque a regressão será aplicada em vários quantis da distribuição de salários, permitindo, dessa maneira, que se investigue como esses diferenciais se comportam em diferentes níveis de renda, ou seja, inclusive nas caudas inferior e superior da distribuição salarial. Isso não seria possível utilizando apenas a regressão linear, porque ela traz a informação dessa diferença na média salarial apenas.

Além de descrever melhor a distribuição da variável de interesse, Koenker (2005) destaca como vantagens da regressão quantílica a equivariância a transformações monótonas e a robustez na presença de outliers e na presença de erros que não seguem a distribuição normal. A primeira tem a ver com a aplicação da função inversa na estimativa do quantil para a variável transformada para estimar o quantil da variável original, já a segunda, com o baixo impacto sofrido por valores extremos nas estimativas do ajuste do modelo.

Por fim, a regressão quantílica se baseia na estimação de parâmetros através da minimização dos erros absolutos, que não permite que o estimador seja calculado diretamente e faz com que sejam necessárias técnicas de programação linear. Dadas essas considerações, não era comum implementar estudos baseados nesse método, mas com a possibilidade de escrever o problema de minimização como um de programação linear esse cenário vem se alterando (SANTOS, 2012).

2 Objetivos

2.1 Objetivo Geral

O objetivo principal deste trabalho é verificar a diferença salarial entre homens e mulheres, com características similares, residentes da área urbana do Distrito Federal no ano de 2018, quantificando essa divergência em diferentes níveis (quantis) de renda.

2.2 Objetivos Específicos

- Estudar regressão quantílica e suas aplicações (metodologia);
- Estudar o impacto de características sociais na renda;
- Comparar perfis de homens e mulheres do Distrito Federal;
- Comparar os resultados da regressão quantílica com os resultados da regressão para a média apresentados no trabalho (ROSA; QUIRINO; XAVIER, 2021).

3 Revisão bibliográfica

3.1 Regressão Linear Múltipla

Segundo Kutner et al. (2005), o modelo de regressão linear pode ser expresso pela equação:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \varepsilon_i,$$

sendo β_0, \dots, β_p os parâmetros; X_{i1}, \dots, X_{ip} são os valores das variáveis X_1, \dots, X_p e ε_i os erros associados ao modelo.

Além do mais, o modelo é construído sob certos pressupostos, sendo eles:

i) $\varepsilon_i \stackrel{iid}{\sim} Normal(0, \sigma^2)$, isto é, os resíduos são independentes e identicamente distribuídos com uma $Normal(0, \sigma^2)$. Logo: a covariância é $cov(\varepsilon_i, \varepsilon_j) = 0$, para todo $i \neq j = 1, \dots, n$, a média é $E(\varepsilon) = 0$ e a variância é $Var(\varepsilon) = \sigma^2$ constante (homocedasticidade).

ii) Y_i são independentes e também seguem uma distribuição Normal com média

$$E(Y_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

e variância

$$Var(Y_i) = \sigma^2.$$

3.1.1 Razão Inversa de Mills

A fim de reduzir a influência de viés de seleção nas estimativas (HECKMAN, 1979), aplicou-se o modelo em duas etapas proposto por Heckman, sendo assim, será incluída a variável Inversa de Mills nas regressões linear e quantílica do salário. Neste trabalho, ela é obtida a partir da probabilidade de um indivíduo estar ou não inserido no mercado de trabalho dado que possui certas características, para isso foi implementada a regressão Probit da variável “Trabalhador” em função de tais fatores (primeira etapa).

Então, considerando as seguintes equações:

$$Y_i = \mathbf{X}_i \beta + \mathbf{U}_{1i} \text{ (regressão de salários) e}$$

$$T_i = \mathbf{W}_i \gamma + \mathbf{U}_{2i} \text{ (regressão de trabalho),}$$

com γ vetor de parâmetros; \mathbf{W}_i vetor com os valores das covariáveis; U_{2i} vetor

de erros; $E(U_{ji}) = 0$ e $E(U_{ji}U_{j'i''}) = \begin{cases} \sigma_{ij'}, & \text{se } i = i'' \\ 0, & i \neq i'' \end{cases}$ (HECKMAN, 1979).

E assumindo que $h(\mathbf{U}_{1i}, \mathbf{U}_{2i})$ é uma densidade Normal Bivariada, Heckman (1979) apresenta

$$E(U_{1i}|U_{2i} \geq -W_i\gamma) = \frac{\sigma_{12}}{\sigma_{22}^{\frac{1}{2}}} \lambda_i,$$

$$E(U_{2i}|U_{2i} \geq -W_i\gamma) = \frac{\sigma_{22}}{\sigma_{22}^{\frac{1}{2}}} \lambda_i,$$

onde

$$\lambda_i = \frac{\phi(Z_i)}{1 - \Phi(Z_i)} = \frac{\phi(Z_i)}{\Phi(-Z_i)},$$

é a razão inversa de Mills, sendo ϕ e Φ , respectivamente, a densidade e a função de distribuição de uma Normal Padrão e $Z_i = -\frac{\mathbf{W}_i\gamma}{\sigma_{22}^{\frac{1}{2}}}$.

3.2 Regressão Quantílica

3.2.1 Quantis

O quantil τ de uma variável corresponde ao valor para o qual $100.\tau\%$ dos valores (ordenados) estão abaixo dele. Por exemplo, suponha que o quantil 0,5 da distribuição de salário-hora seja R\$ 15,00, isso quer dizer que 50% dos salários analisados variam até R\$ 15,00.

Temos também que, como uma variável aleatória pode ser caracterizada por sua função de distribuição (continua à direita)

$$F(X) = P(X \leq x),$$

o τ -ésimo quantil da distribuição de X pode ser obtido por

$$F^{-1}(\tau) = \inf\{x : F(x) \geq \tau\}, 0 < \tau < 1.$$

Mas, no contexto de regressão quantílica, é interessante definir o quantil a partir da função de perda esperada (KOENKER, 2005), isto é, encontrando o \hat{x} que minimize

$$\rho_\tau(u) = u(\tau - I(u < 0)), 0 < \tau < 1.$$

Então,

$$E[\rho_\tau(X - \hat{x})] = (\tau - 1) \int_{-\infty}^{\hat{x}} (x - \hat{x})dF(x) + \tau \int_{\hat{x}}^{\infty} (x - \hat{x})dF(x),$$

diferenciando em x e igualando a zero, temos

$$0 = (1 - \tau) \int_{-\infty}^{\hat{x}} dF(x) - \tau \int_{\hat{x}}^{\infty} dF(x) = F(\hat{x}) - \tau.$$

Como F é monótona, qualquer elemento em $\{x : F(x) = \tau\}$ minimiza a perda esperada. Quando a solução for única, $\hat{x} = F^{-1}(\tau)$ será o τ -ésimo quantil que minimiza $\rho(u)$, caso contrário, teremos um intervalo de τ -ésimos quantis, dos quais devemos selecionar o menor para que a função quantílica empírica seja contínua à esquerda (KOENKER, 2005).

Se $\tau = 1/2$,

$$E[\rho_{1/2}(X - \hat{x})] = (1/2 - 1) \int_{-\infty}^{\hat{x}} (x - \hat{x})dF(x) + 1/2 \int_{\hat{x}}^{\infty} (x - \hat{x})dF(x) = 1/2E[|X - \hat{x}|],$$

então minimizar $E[\rho_{1/2}(X - \hat{x})]$ equivale a minimizar $E[|X - \hat{x}|]$, que resulta em \hat{x} igual ao quantil de ordem $1/2$, ou seja, a mediana.

Essa definição nos permite pensar em modelos de estimação de quantis condicionais. Bom, sabemos que a média amostral de uma variável Y soluciona o problema

$$\min_{\mu \in \mathbb{R}} \sum_{i=1}^n (y_i - \mu)^2.$$

Logo, se desejamos expressar a média condicional de y dado \mathbf{x} como $\mu(x) = \mathbf{x}^T \beta$, β pode ser estimado por mínimos quadrados, resolvendo

$$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n (y_i - \mathbf{x}_i^T \beta)^2.$$

Analogamente, Koenker e Bassett (1978) mostraram que o quantil amostral de ordem τ , $\hat{\alpha}(\tau)$, resolve

$$\min_{\alpha \in \mathbb{R}} \sum_{i=1}^n \rho_\tau(y_i - \alpha)$$

e podemos especificar o τ -ésimo quantil condicional de Y dado \mathbf{x} como $Q_y(\tau|x) = \mathbf{x}\beta(\tau)$ e encontrar $\hat{\beta}(\tau)$ que solucione

$$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \rho_{\tau}(y_i - \mathbf{x}_i^T \beta).$$

Assim, se no método de mínimos quadrados, descrevemos a relação entre X e Y como

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i,$$

em que ϵ_i com média 0, dizemos que a média condicional de Y dado X pode ser escrita da seguinte maneira

$$E(Y|X) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p.$$

Para estudar os quantis condicionais de Y dado X, supondo relações lineares do tipo

$$y_i = \beta_0(\tau) + \beta_1(\tau)x_{i1} + \cdots + \beta_p(\tau)x_{ip} + u_i,$$

sendo u_i variáveis aleatórias independentes e identicamente distribuídas com quantil de ordem τ igual a 0, o quantil condicional τ de Y dado X é dado por

$$Q_{\tau}(Y|x) = \beta_0(\tau) + \beta_1(\tau)x_1 + \cdots + \beta_p(\tau)x_p.$$

Nesse caso, o vetor de parâmetros β é indexado a τ porque um dos objetivos é verificar se os parâmetros mudam em função de τ 's diferentes.

Se as estimativas $\hat{\beta}(\tau)$ são muito próximas para diferentes valores de (τ) , há evidência de que os erros são independentes e identicamente distribuídos. Caso contrário, isto é, se os coeficientes estimados variam de acordo com (τ) , os erros podem apresentar algum tipo de heterocedasticidade, o que significa que modelos de regressão quantílica podem incorporar essa possível diferença de variância.

3.2.2 Estimadores

Considere o seguinte modelo

$$\mathbf{Y} = \mathbf{X}\beta(\tau) + \epsilon,$$

no qual \mathbf{Y} é um vetor de observações $n \times 1$, \mathbf{X} é uma matriz de constantes conhecidas $n \times p$, $\beta(\tau)$ é um vetor $p \times 1$ de parâmetros desconhecidos e ϵ é um vetor de erros independentes e identicamente distribuídos com função de distribuição F e quantil de ordem (τ) igual a zero.

Como mencionado na seção anterior, $\hat{\beta}(\tau)$ é obtido através do resultado de $\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \rho_{\tau}(y_i - \mathbf{x}_i^T \beta)$, mas Koenker (2005) mostra que essa minimização pode ser escrita como um programa linear da forma

$$\min_{(\beta, \mathbf{u}, \mathbf{v}) \in \mathbb{R}^p \times \mathbb{R}_+^{2n}} \{ \tau \mathbf{I}_n^T \mathbf{u} + (1 - \tau) \mathbf{I}_n^T \mathbf{v} \mid \mathbf{X} \beta + \mathbf{u} - \mathbf{v} = \mathbf{y} \},$$

na qual \mathbf{X} representa a matriz de regressão $n \times p$, \mathbf{I}_n^T um vetor $1 \times n$ de 1's, \mathbf{u} e \mathbf{v} são vetores $n \times 1$, sendo u_i e v_i seus termos, respectivamente, definidos como

$$u_i = \begin{cases} y_i - \hat{y}_i, & \text{se } y_i - \hat{y}_i > 0 \\ 0, & \text{caso contrário} \end{cases}$$

$$v_i = \begin{cases} y_i - \hat{y}_i, & \text{se } y_i - \hat{y}_i < 0 \\ 0, & \text{caso contrário} \end{cases}$$

com $\hat{y}_i = \mathbf{x}_i^T \hat{\beta}$.

3.2.3 Intervalos de Confiança

Em (KOENKER, 2005) são apresentados três métodos para a estimação dos intervalos de confiança para modelos de regressão quantílica, métodos baseados em resultados assintóticos, bootstrap e testes de escores ordinais. A princípio focaremos no primeiro, já que trabalharemos com uma amostra relativamente grande e os outros métodos exigiriam um maior esforço computacional por conta disso.

Método baseado em resultados assintóticos

Diferentemente da regressão linear, na qual conseguimos obter um estimador fechado para o vetor de parâmetros β e construir intervalos de confiança exatos, aqui trabalharemos com resultados assintóticos. Podemos estimar matriz de covariâncias assintótica do vetor $\hat{\beta}$ de duas formas diferentes, como descrito a seguir.

Koenker e Bassett (1978) citado em (SANTOS, 2012) provaram o seguinte teorema.

Teorema 1. Seja $\{\hat{\beta}(\tau_1), \hat{\beta}(\tau_2), \dots, \hat{\beta}(\tau_m)\}$, com $0 < \tau_1 < \tau_2 < \dots < \tau_m < 1$ uma sequência de estimadores para os parâmetros do modelo $\mathbf{Y} = \mathbf{X}\beta(\tau) + \epsilon$. Seja $\xi_i(\tau_i) = F^{-1}(\tau_i)$ o quantil de ordem τ_i e assuma que

- (i) F é contínua e tem densidade contínua e positiva em $\xi_i(\tau_i)$, para $i = 1, 2, \dots, m$.
- (ii) A matriz \mathbf{X} de planejamento tem uma coluna de uns.
- (iii) $\lim_{n \rightarrow \infty} n^{-1} \mathbf{X}' \mathbf{X} = \mathbf{Q}$ matriz positiva definida.

Nessas condições,

$$\sqrt{n}(\hat{\beta}(\tau_1) - \beta(\tau_1), \dots, \hat{\beta}(\tau_m) - \beta(\tau_m)) \xrightarrow{D} N_{m \times p}(\mathbf{0}, V(\tau_1, \dots, \tau_m)),$$

em que a matriz de covariâncias, $V(\tau_1, \dots, \tau_m)$, pode ser definida como

$$\Omega(\tau_1, \dots, \tau_m; F) \otimes Q^{-1},$$

sendo $\Omega(\tau_1, \dots, \tau_m; F)$ a matriz de covariâncias entre m quantis amostrais de amostras aleatórias com distribuição F e \otimes indica o produto de Kronecker.

Se apenas um quantil τ for de interesse, segundo Kocherginsky et al. (2005), citado em (SANTOS, 2012), a matriz de covariâncias assintótica de $\hat{\beta}(\tau)$ quando os erros não são identicamente distribuídos é

$$V(\tau) = \tau(1 - \tau)(X'FX)^{-1}(X'X)(X'FX)^{-1},$$

em que $F = (f_1(0), \dots, f_n(0))$ é a matriz diagonal e $f_j, j = 1, \dots, n$, a função densidade dos erros.

Para estimá-la, pode-se substituir o $f_i(0)$ na matriz $X'FX$ por uma estimativa assintoticamente não viciada, por exemplo, por

$$\frac{2h_n}{\mathbf{x}'_1 \hat{\beta}_{\tau+h_n} - \mathbf{x}'_1 \hat{\beta}_{\tau-h_n}},$$

implementada no pacote "quantreg" do software R (SANTOS, 2012).

Mas se os erros são identicamente distribuídos, quando $f_1(x) = \dots = f_n(x) = f(x)$, ela se torna

$$V(\tau) = \frac{\tau(1 - \tau)}{f^2(0)}(X'X)^{-1}$$

e neste caso, de erros com mesma distribuição, segundo Kocherginsky et al. (2005), citado em (SANTOS, 2012), é possível estimar $1/f(0)$ usando uma diferença de quantis empíricos dos resíduos, com

$$\frac{\hat{F}^{-1}(\tau + h_n) - \hat{F}^{-1}(\tau - h_n)}{2h_n}$$

em que $\lim_{n \rightarrow \infty} h_n = 0$.

A partir dos resultados do Teorema 1 e da estimativa da matriz $V(\tau)$, os intervalos de confiança para os parâmetros $\beta_i(\tau_i)$ podem ser obtidos.

4 Materiais e métodos

4.1 Materiais

Os dados utilizados no presente trabalho foram obtidos pela Pesquisa Distrital por Amostra de Domicílios 2018 (PDAD 2018), fornecidos pela Companhia de Planejamento do Distrito Federal, em duas planilhas separadas: uma se refere aos moradores e possui 69.654 observações de 103 variáveis; a outra é sobre os domicílios e apresenta 21.908 observações de 130 variáveis.

A PDAD é uma pesquisa regulamentada pelo Decreto Nº 39.403, de 26 de outubro de 2018, que teve início em 2004 e é realizada de 2 em 2 anos. Além disso, ela é representativa para as trinta e uma regiões administrativas (RA's) do Distrito Federal e tem como objetivo principal reconhecer as características socioeconômicas e de migração da população urbana da região a fim de subsidiar a formulação ou reestruturação de políticas públicas que mais se adequem e melhor atendam os habitantes da localidade em questão (ROSA; QUIRINO; XAVIER, 2021).

Na pesquisa são levantadas questões sobre características gerais dos moradores e dos domicílios, entre esses aspectos estão a idade, o sexo, a raça, a escolaridade, o local de trabalho e moradia e o rendimento, característica fundamental para este estudo.

4.1.1 Variáveis

Considerando aspectos relevantes para a determinação salarial, serão abordadas as seguintes variáveis:

Quadro 1: Variáveis obtidas a partir da PDAD 2018, Distrito Federal - 2018

Variável	Descrição
Salário-hora	valor recebido, em reais, por hora trabalhada no trabalho principal
Trabalhador	identifica se a pessoa está trabalhando
Idade	idade da pessoa
Mulher	variável de interesse, identifica se a pessoa é do sexo feminino
Negro	identifica se a pessoa é preta/parda
Deficientes	identifica se a pessoa tem alguma deficiência auditiva, visual, motora ou intelectual grave
Casado	identifica se a pessoa está casada
Estudante	identifica se a pessoa frequenta alguma instituição de ensino
Ensino Médio	identifica se a pessoa tem ensino médio completo
Ensino Superior	identifica se a pessoa tem ensino superior completo
Filhos no domicílio	identifica se no domicílio existia algum filho
Idosos no domicílio	identifica se no domicílio existia alguém com 60 anos ou mais
Quantidade de trabalhos	identifica se a pessoa possuía um trabalho ou mais
Função trabalho	identifica atividade desempenhada pela pessoa no trabalho
Setor de trabalho	identifica atividade principal da empresa onde trabalha
Setor informal	identifica se a pessoa trabalha no setor informal
Renda dom. s/ renda indivíduo	logaritmo da renda domiciliar excluindo a renda da pessoa
Setor da economia	identifica a qual setor da economia a empresa onde a pessoa trabalha pertence
Setor informal	identifica se a pessoa trabalha no setor informal
Setor público	identifica se a pessoa trabalha no setor público
Tempo no emprego	tempo, em anos, que a pessoa estava no trabalho principal
Trabalha no Plano Piloto	identifica se a pessoa trabalha no Plano Piloto
Grupos de RA	agrupa as regiões administrativas de moradia, segundo a PED
Inversa de Mills	probabilidade de estar trabalhando segundo características pessoais

¹ Fonte: Elaboração própria, dados da PDAD 2018 - Codeplan.

4.2 Métodos

Com respeito à análise dos dados, serão empregadas as técnicas de regressão linear e quantílica a fim de explicar o salário-hora com base nas características de sexo, raça, deficiência, idade, escolaridade, área de trabalho e locais de trabalho e moradia.

Tendo em vista a assimetria da variável salário-hora o $\log(\text{salário-hora})$ configura-se como variável resposta e as demais características como variáveis explicativas em ambas as técnicas, semelhante ao procedimento realizado em (ROSA; QUIRINO; XAVIER, 2021).

Portanto, o modelo proposto na regressão linear será do tipo:

$$\log(Y_i) = \beta_0 + \mathbf{X}\beta + \varepsilon,$$

¹PED=Pesquisa de Emprego e Desemprego.

sendo $\log(Y)$ o vetor de $\log(\text{salários-hora})$, β_0 o intercepto, \mathbf{X} é uma matriz de características dos trabalhadores, β é um vetor de coeficientes para essas características e ϵ os erros associados ao modelo.

Enquanto o modelo proposto na regressão quantílica será do tipo:

$$Q_\tau(\mathbf{\log(S)}) = \beta_0(\tau) + \mathbf{X}\beta(\tau) + \epsilon,$$

no qual $Q_\tau(\mathbf{\log(S)})$ é o τ -ésimo quantil do vetor de $\log(\text{salários-hora})$, $\beta_0(\tau)$ é o intercepto, \mathbf{X} é uma matriz de características dos trabalhadores, $\beta(\tau)$ é um vetor de coeficientes para essas características e ϵ é um vetor de erros independentes e identicamente distribuídos com função de distribuição F e quantil de ordem (τ) igual a zero.

Os resultados obtidos para diferentes quantis serão comparados com os do trabalho (ROSA; QUIRINO; XAVIER, 2021), que se apoiou apenas na regressão para a média.

5 Resultados

5.1 Análise Exploratória

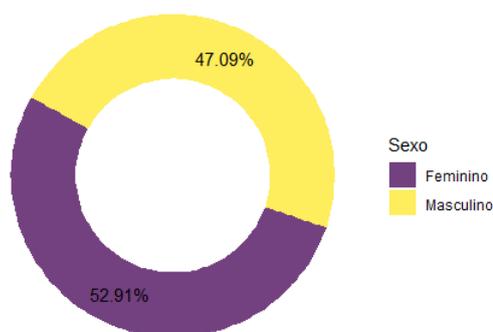
A análise exploratória permite uma visão inicial dos dados analisados e um melhor entendimento de suas principais características. Desta forma, é possível escolher o método de análise mais adequado para aquele conjunto, bem como formular hipóteses de estudo.

Nesta seção, serão descritas as principais características dos indivíduos presentes na amostra e o perfil dos salários.

5.1.1 Características dos indivíduos

A amostra pesquisou um total de 69.654 pessoas de 21.908 domicílios localizados na área urbana do Distrito Federal em 2018. Dessas, 36.855 são mulheres e 32.799 homens, essa relação é abordada no Gráfico 1. Esse resultado condiz com a realidade da população brasileira, que apresenta ligeiramente mais mulheres que homens.

Gráfico 1: Proporção de homens e mulheres, Distrito Federal - 2018

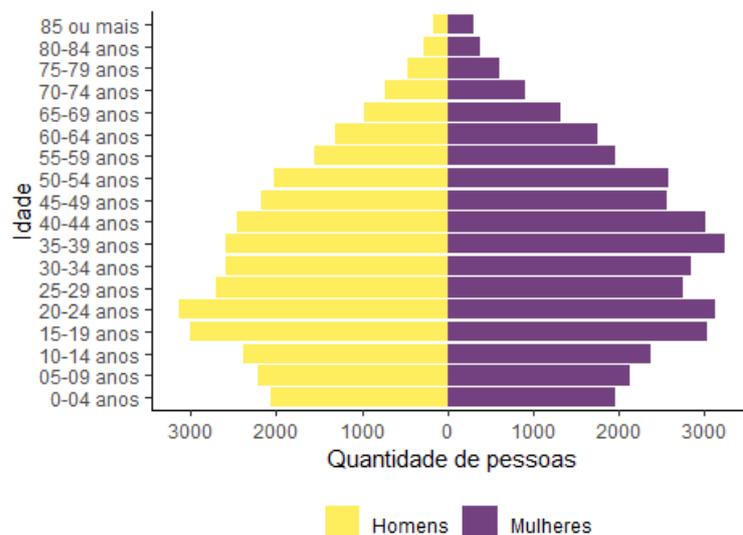


Abaixo, há a caracterização dos moradores em função do sexo e idade.

A partir da pirâmide etária, ilustrada no Gráfico 2, percebe-se, no ano de 2018, que o Distrito Federal apresenta uma grande concentração de pessoas com idades entre 15 e 54 anos, revelando uma população adulta.

Também é interessante notar que, conforme as idades aumentam, há um indício de transposição do padrão de mais mulheres que homens em cada faixa etária, já que nas idades mais jovens a quantidade de homens supera a de mulheres e a partir de 30 anos ocorre o contrário.

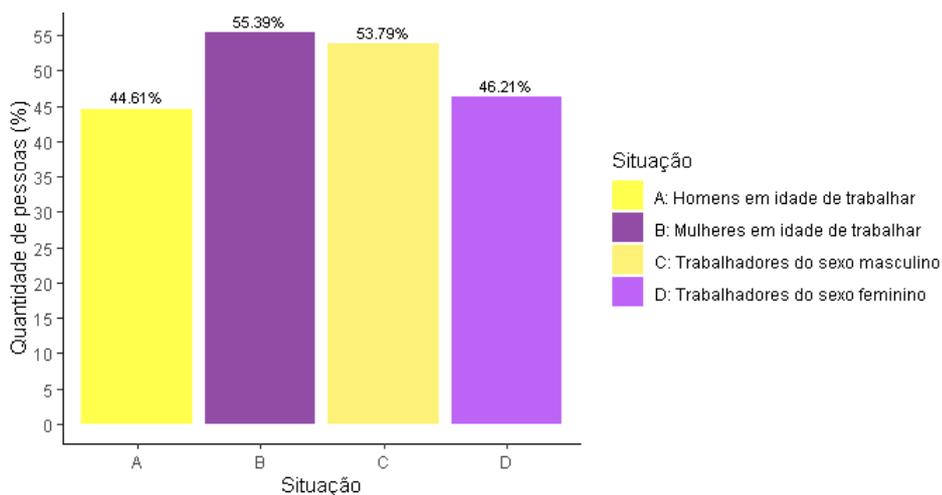
Gráfico 2: Pirâmide etária, Distrito Federal - 2018



A seguir, serão analisadas características dos moradores em idade de trabalhar, ou seja, com, no mínimo, 14 anos de idade².

Em se tratando de mercado de trabalho³, o Gráfico 3 aborda uma relação entre a proporção de pessoas que podem trabalhar e a de pessoas que efetivamente possuem algum trabalho, segundo sexo.

Gráfico 3: Pessoas em idade de trabalhar x Pessoas que trabalham, segundo sexo, Distrito Federal - 2018



Do Gráfico 3, percebe-se que apesar de a proporção de mulheres acima de 14 anos ser maior que a de homens, a proporção de mulheres que trabalham é menor, uma vez

²Retirado de: <https://www.ibge.gov.br/explica/desemprego.php>

³Desconsidera pessoas com menos de 14 anos e que não receberam pelo trabalho principal.

que elas correspondem a 55,39% do total de pessoas em idade de trabalhar e a 46,21% do total de trabalhadores. Sendo assim, é interessante calcular as taxas de desocupação segundo sexo, conforme apresentado na Tabela 1.

Tabela 1: Taxas de desocupação segundo sexo, Distrito Federal - 2018

Sexo	Taxa de desocupação(%)
Ambos	13,64
Feminino	16,19
Masculino	11,39

Segundo o Instituto Brasileiro de Geografia e Estatística - IBGE, a força de trabalho é composta por pessoas com mais de 14 anos que estão trabalhando ou procurando trabalho e a taxa de desocupação é a porcentagem de pessoas desempregadas na força de trabalho⁴. Conforme a Tabela 1, a taxa de desemprego no DF é 13,64%, enquanto a dos homens é 11,39% e a das mulheres, 16,19%.

A Tabela 2 traz informações sobre a idade e o tempo no emprego dos cidadãos. Observa-se que 50% das pessoas em idade de trabalhar têm até 38 anos e 50% dos que trabalham têm até 4 anos de prestação de serviço na mesma empresa.

Tabela 2: Medidas descritivas de idade e tempo no emprego, Distrito Federal - 2018

Variável	Mín	Q(0.25)	Mediana	Média	Q(0.75)	Máx
Idade	14	24	38	40.4	54	104
Tempo no emprego	0	1.5	4	7.79	10	41.67

No que diz respeito à autodeclaração de raça dos indivíduos em 2018, observamos o cenário ilustrado no Gráfico 4:

De acordo com o Gráfico 4, nota-se que a maior parte da população é composta por pardos, brancos e negros, nessa ordem. Além disso, a proporção de mulheres amarelas, brancas e indígenas é maior que a de homens, enquanto que a de mulheres pardas e negras, menor.

A Tabela 3 traz a proporção de pessoas segundo grau de limitação em alguma das seguintes faculdades: de enxergar; de ouvir; de caminhar ou subir degraus e mental.

⁴Retirado de: <https://www.ibge.gov.br/explica/desemprego.php>

Gráfico 4: Raça segundo sexo, Distrito Federal - 2018

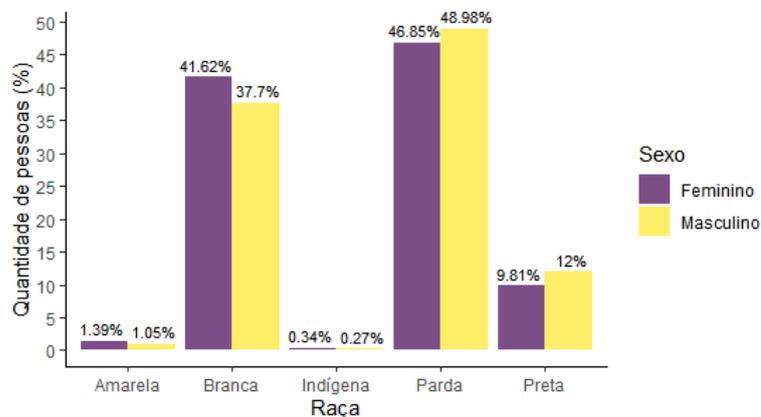
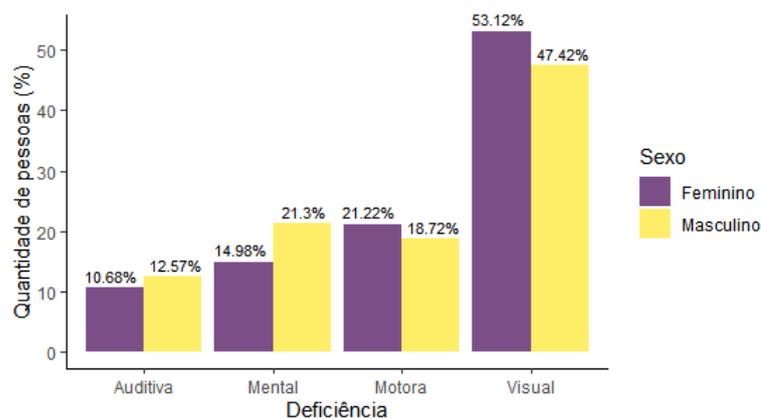


Tabela 3: Portadores de deficiência segundo sexo, Distrito Federal - 2018

Dificuldade	Sexo	
	Feminino (%)	Masculino (%)
Alta/Total	4,27	2,96
Baixa/Nenhuma	51,12	41,65

A proporção de pessoas com limitações graves é baixa em relação ao total de trabalhadores (7.23%), mas é maior entre as pessoas do sexo feminino. O Gráfico 5 apresenta uma caracterização dos que possuem dificuldade alta ou total nas faculdades citadas acima.

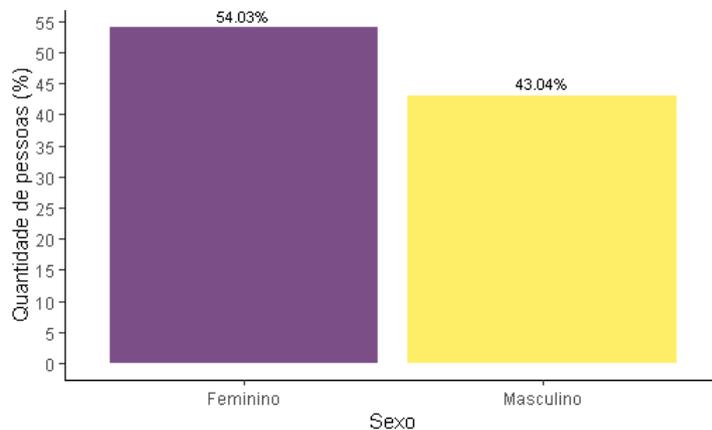
Gráfico 5: Portadores de deficiência grave segundo sexo, Distrito Federal - 2018



A deficiência visual e a motora foram as mais reportadas entre o total de pessoas acima de 14 anos e entre as mulheres, enquanto auditiva e mental foram mais apontadas por homens.

O Gráfico 6 informa a proporção de pessoas com mais de 14 anos que residem ou não com filhos segundo sexo.

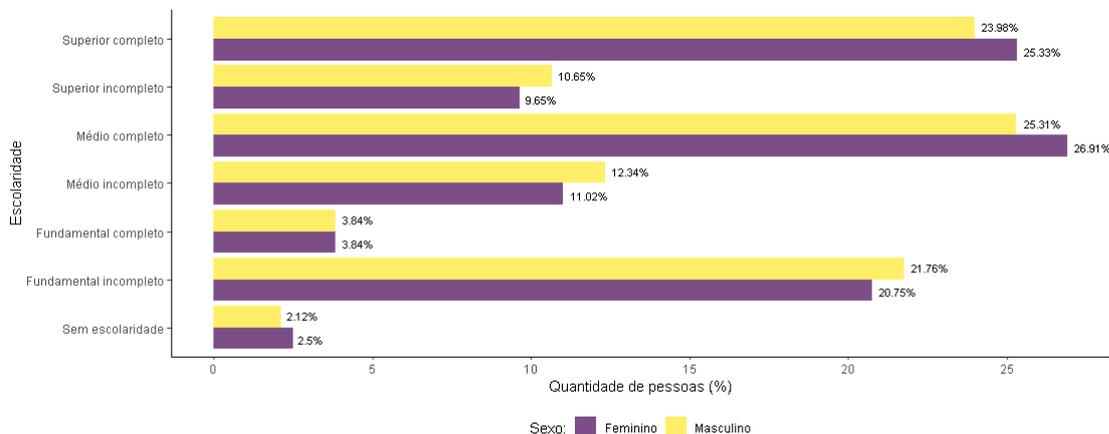
Gráfico 6: Residir com filhos segundo sexo, Distrito Federal - 2018



Assim como ilustrado na figura 6, a disposição de filhos é bem diferente entre os sexos, sendo a proporção de mães que residem com filhos maior do que a de pais nessa mesma condição.

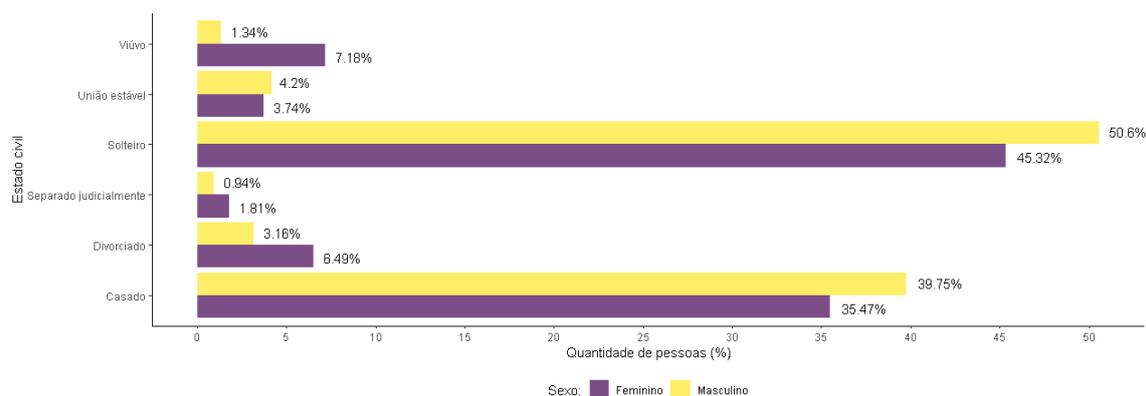
A respeito do grau de escolaridade segundo sexo, temos a figura 7.

Gráfico 7: Nível de escolaridade segundo sexo, Distrito Federal - 2018



Conforme ilustrado na figura 7, a proporção de pessoas do sexo feminino sem escolaridade, que cursaram ensino médio e superior completos é maior que a do masculino. Por outro lado, a proporção de homens que não completaram o ensino fundamental, o médio e o superior ultrapassa a de mulheres. O que pode indicar uma maior evasão escolar de pessoas do sexo masculino.

Gráfico 8: Estado civil segundo, Distrito Federal - 2018



De acordo com a figura 8, os estados civis mais comuns entre pessoas acima de 14 anos no Distrito Federal, em 2018, são “solteiro” e “casado”, sendo a proporção de homens maior que a de mulheres em ambas categorias. A proporção de mulheres viúvas e de divorciadas supera a de homens.

Conforme as Tabelas 4 e 5, que informam a posição ocupada no trabalho (cargo) e a área de trabalho (atividade da empresa) no emprego principal, de acordo com o sexo do trabalhador, a maior parte das mulheres trabalha na posição de: empregado (64,89%); conta própria ou autônomo (16,49%) e empregado doméstico (9,97%). E nos setores de: Outras Atividades De Serviços (21,75%); Comércio, Reparação De Veículos Automotores E Motocicletas (16,66%); Serviços Domésticos (14,90%); Educação (11,42%) e Saúde Humana E Serviços Sociais (9,83%). Já os homens ocupam cargos de: empregado (62,47%); conta própria ou autônomo (24,76%); Militar do Exército, Marinha ou Aeronáutica (2,96%) e Polícia militar, Civil, Bombeiros (2,37%). E nos setores de: Outras Atividades De Serviços (28,71%); Comércio, Reparação De Veículos Automotores E Motocicletas (17,67%); Administração Pública, Defesa E Seguridade Social (10,07%); Construção (8,16%) e Transporte, Armazenagem E Correio (5,97%).

Isso posto, as diferenças que mais chamam atenção entre os sexos são dos trabalhadores autônomos, com maior expressividade do sexo masculino (24,76%, contra 16,49% das mulheres), e dos trabalhadores domésticos com maior expressividade do sexo feminino (9,97% contra 0,3% dos homens). Comparativamente, os homens também se apresentam mais nas áreas mais braçais da construção, transporte, armazenagem e indústria, enquanto que as mulheres tem maior participação relativa nas áreas de educação, saúde e serviços domésticos. Esse fato remonta à ideia de existência de segregação ocupacional em função do sexo.

Tabela 4: Posição no trabalho segundo sexo, Distrito Federal - 2018

Posição no Trabalho	Sexo	
	Feminino (%)	Masculino (%)
Aprendiz	0.68	0.56
Cargo comissionado no setor público	1.07	0.85
Conta Própria ou Autônomo	16.49	24.76
Dono de Negócio Familiar	0.59	0.65
Empregado (Exceto Empregado Doméstico)	64.89	62.47
Empregado Doméstico	9.97	0.3
Empregado Temporário	0.75	0.48
Empregador	0.78	1.64
Estágio Remunerado	2.7	1.9
Militar do Exército, Marinha ou Aeronáutica	0.51	2.96
Polícia Militar, Polícia Civil, Corpo de Bombeiros e	0.68	2.37
Presta Serviço Militar Obrigatório	-	0.09
Profissional Universitário Autônomo (Profissional Liberal)	0.67	0.71
Religioso remunerado (padre, pastor e outros)	0.01	0.09
Sócio de Cooperativa	0.18	0.15
Trabalhador familiar sem remuneração salarial	0.03	0.01

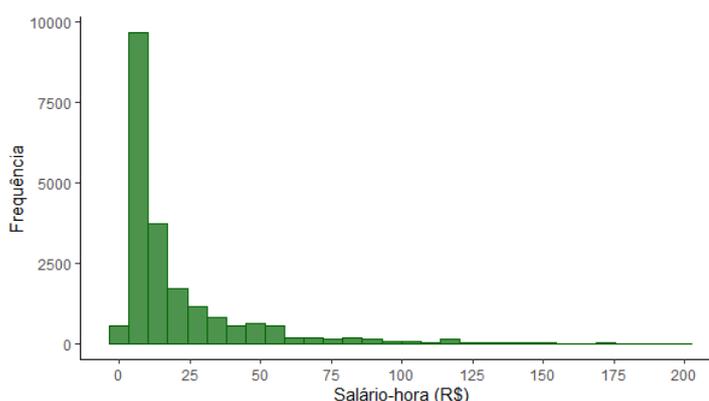
Tabela 5: Área de trabalho segundo sexo, Distrito Federal - 2018

Área de Trabalho	Sexo	
	Feminino (%)	Masculino (%)
Administração Pública, Defesa E Seguridade Social	6.82	10.07
Agricultura, Pecuária, Produção Florestal, Pesca E Aquicultura	0.23	0.89
Água, Esgoto, Atividades De Gestão De Resíduos E Descontaminação	0.06	0.32
Alojamento E Alimentação	2.39	2.36
Artes, Cultura, Esporte E Recreação	1.21	1.18
Atividades Administrativas E Serviços Complementares	7.39	4.86
Atividades Financeiras, De Seguros E Serviços Relacionados	1.9	2.14
Atividades Imobiliárias	0.32	0.42
Atividades Profissionais, Científicas E Técnicas	1.54	2.45
Comércio, Reparação De Veículos Automotores E Motocicletas	16.66	17.67
Construção	0.43	8.16
Educação	11.42	3.95
Eletricidade E Gás	0.04	0.98
Indústrias De Transformação	0.22	0.99
Indústrias Extrativas	0.08	0.47
Informação E Comunicação	1.97	3.94
Organismos Internacionais E Outras Instituições Extraterritoriais	0.05	0.07
Outras Atividades De Serviços	21.75	28.71
Saúde Humana E Serviços Sociais	9.83	3
Serviços Domésticos	14.9	1.4
Transporte, Armazenagem E Correio	0.8	5.97

5.1.2 Salários

Agora serão analisadas as questões de rendimento em função das características abordadas. Nos Gráficos seguintes o valor do salário-hora foi limitado a R\$ 200,00 para que a visualização não fosse comprometida. O histograma abaixo mostra a distribuição do salário recebido por hora trabalhada.

Gráfico 9: Distribuição de salários-hora, Distrito Federal - 2018



Pelo Gráfico 9 é possível ver que a distribuição é assimétrica à direita, já que os salários estão concentrados nos menores valores, neste caso, abaixo de R\$25,00. Isso era de se esperar, uma vez que a maioria das pessoas apresenta salários baixos, enquanto poucas delas apresentam os maiores salários.

A Tabela 6 a seguir traz informações sobre o salário, por hora, dos trabalhadores.

Tabela 6: Medidas descritivas do salário-hora (em reais), Distrito Federal - 2018

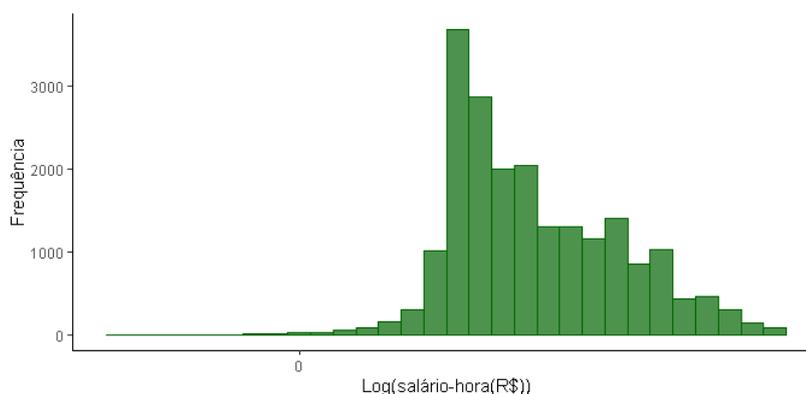
Mín	Q(0.1)	Q(0.25)	Mediana	Média	Q(0.75)	Q(0.9)	Máx
0,16	5,08	6,27	10,45	22,14	24,71	51,6	2272,17

As medidas condizem com o histograma ilustrado na figura 9. Também é possível ter uma melhor noção do grau da assimetria da distribuição, já que 75% dos trabalhadores têm salário até R\$24,71 e 90% deles recebem no máximo R\$51,60, enquanto alguns valores ultrapassam R\$200,00 e chegam a R\$2272,17.

Tendo isso em vista e a fim de aplicar a regressão linear, que necessita de alguns pressupostos, será utilizada a transformação logarítmica do salário-hora, cuja distribuição é apresentada no Gráfico 10.

Dado o gráfico 10, percebe-se que a assimetria de log (salário-hora) é menos pro-

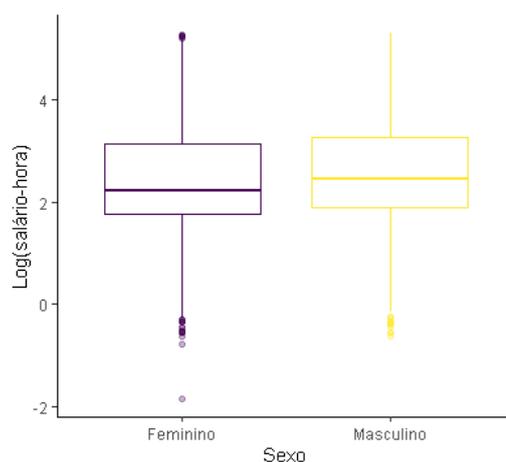
Gráfico 10: Distribuição de salários-hora, Distrito Federal - 2018



nunciada que a de salário-hora.

O gráfico 11 mostra a distribuição de salários segundo sexo.

Gráfico 11: Distribuição de salários-hora, segundo sexo, Distrito Federal - 2018



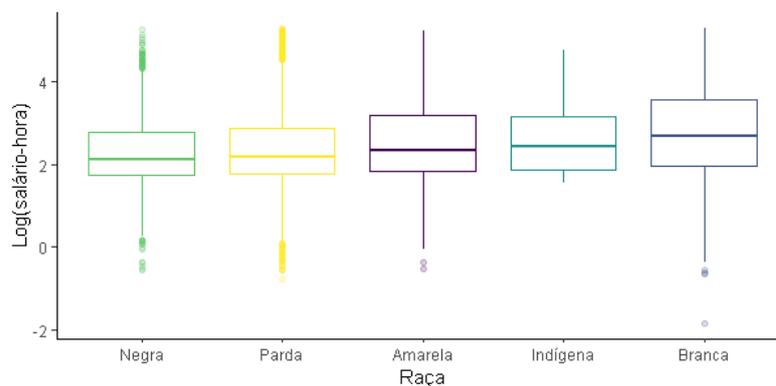
Analisando a distribuição geral de salários, percebe-se, pelos boxplots apresentados na figura 11, que ambos os sexos possuem rendas fora do comum (valores representados pelos pontos abaixo das hastes das caixas de boxplot). Ademais, o terceiro quartil para os homens está um pouco acima do valor das mulheres, mas não se nota uma grande diferença entre as remunerações obtidas por cada sexo, por conta dos valores próximos.

A distribuição de salário segundo raça é retratada na figura 12:

Como as distribuições apresentam certas diferenças em função da raça, essa pode ser uma variável relevante para o estudo, conforme indicado na figura 12.

Observa-se que as distribuições de negros e pardos são bem parecidas e, além disso, apresentam menores medianas, então pode-se pensar em juntar tais categorias.

Gráfico 12: Distribuição de salários-hora, segundo raça, Distrito Federal - 2018

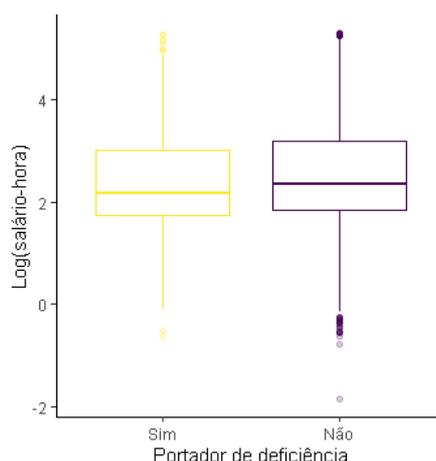


Pessoas brancas apresentam os maiores salários, já que sua mediana supera as das outras características, assim como o terceiro quartil está a uma diferença razoável dos demais.

Por fim, o fato de a proporção de pessoas amarelas e a de indígenas serem baixas (havendo, assim, menos informação e mais incerteza sobre elas) pode ter produzido maiores estimativas para os salários.

O Gráfico 13 trata do salário em função de ser portador de alguma deficiência. Bem como no caso do sexo, ao olhar para a distribuição geral segundo ser PCD (pessoa

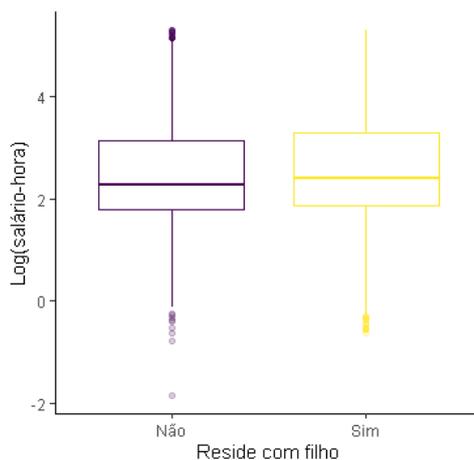
Gráfico 13: Distribuição de salários-hora, segundo porte de deficiência, Distrito Federal - 2018



com deficiência), não parece haver uma diferença significativa dos salários, contudo podemos observar que aqueles que não são PCD apresentam salários um pouco maiores e a quantidade de valores altos e atípicos para os que são PCD, também parece ser menor.

O Gráfico 14 aborda o rendimento salarial em consequência de ter filhos.

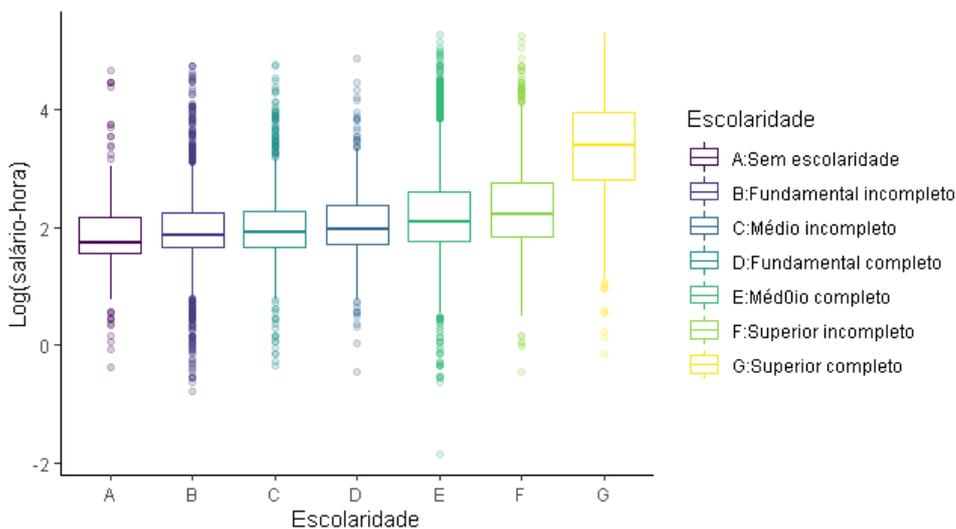
Gráfico 14: Distribuição de salários-hora, segundo residência com filhos, Distrito Federal - 2018



A partir do Gráfico 14, não se verifica uma diferença considerável entre os rendimentos de quem mora e de quem não mora com filhos, apesar de os salários de quem reside com filhos serem um pouco maiores do que os de quem não reside.

O Gráfico 15 mostra o salário em função do nível de escolaridade do trabalhador.

Gráfico 15: Distribuição de salários-hora, segundo nível de escolaridade, Distrito Federal - 2018

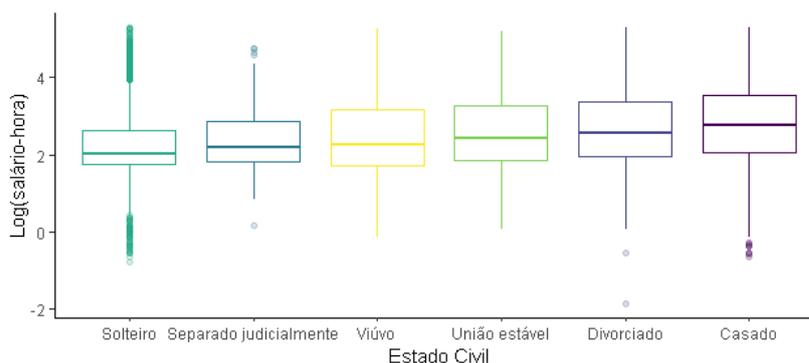


Com base no Gráfico 15, percebe-se que a remuneração obtida cresce, conforme se completa um nível de instrução maior. Isso pode estar atrelado ao fato de que atividades que requerem maior grau de especialização tendem a apresentar menor oferta de mão de obra do que aquelas que requerem um menor grau.

A figura 16 apresenta o salário em função do estado civil do indivíduo.

Conforme o Gráfico 8, as pessoas estão divididas principalmente entre solteiras e

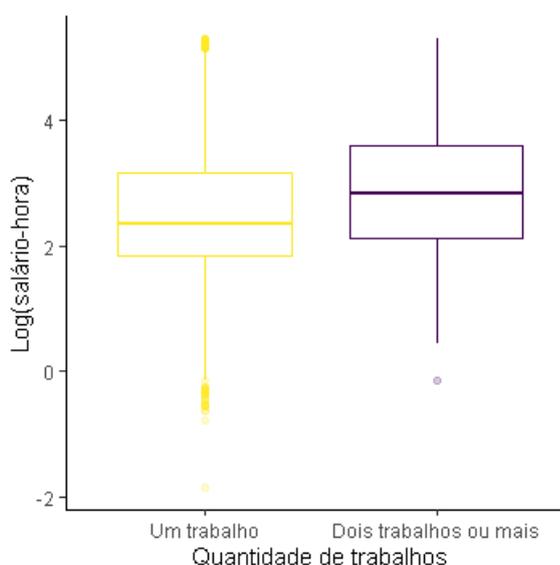
Gráfico 16: Distribuição de salários-hora, segundo estado civil, Distrito Federal - 2018



casadas, já a partir do Gráfico 16, nota-se que indivíduos com esses atributos apresentam os menores e os maiores salários, respectivamente, havendo uma diferença relevante entre as distribuições salariais de solteiros e casados.

Outra questão interessante é a quantidade de trabalhos que o indivíduo possui. Essa variável é apresentada na figura 17 a seguir.

Gráfico 17: Distribuição de salários-hora, segundo quantidade de trabalhos, Distrito Federal - 2018



De acordo com a figura 17, a renda por hora do trabalho principal, parece ter relação com a quantidade de empregos, de forma que, quem tem mais de um emprego, também apresenta maior mediana e quantil 0,75 de salário.

O cargo no trabalho é uma questão que pode ter influência na remuneração obtida, sendo assim, os Gráficos 18 e 19 abordam essa variável. A fim de obter uma melhor visualização dos dados, em virtude da grande quantidade de categorias presentes, as informações foram divididas nos Gráficos 18 e 19, com o eixo x organizado em ordem

crecente das medianas dos salários.

Gráfico 18: Distribuição de salários-hora, segundo função no trabalho, Distrito Federal - 2018

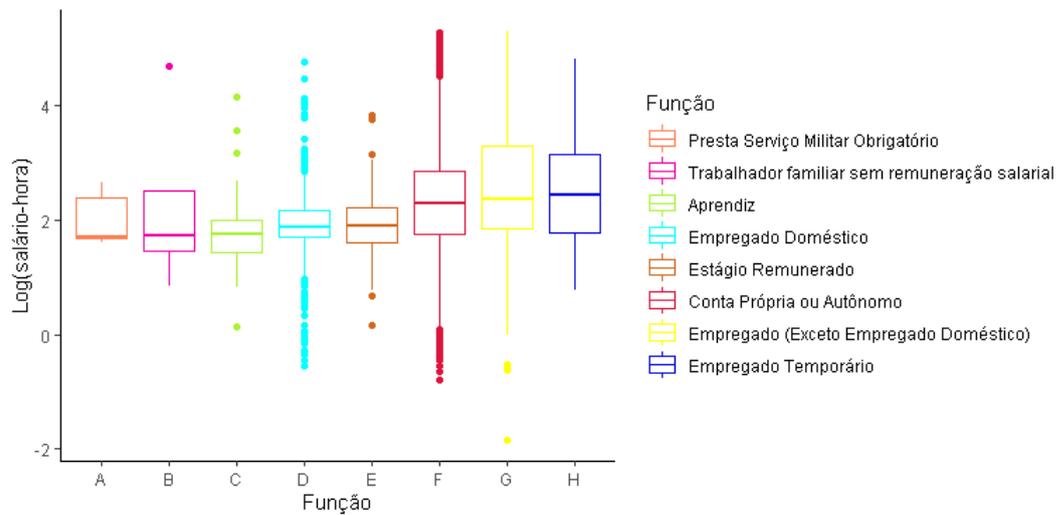
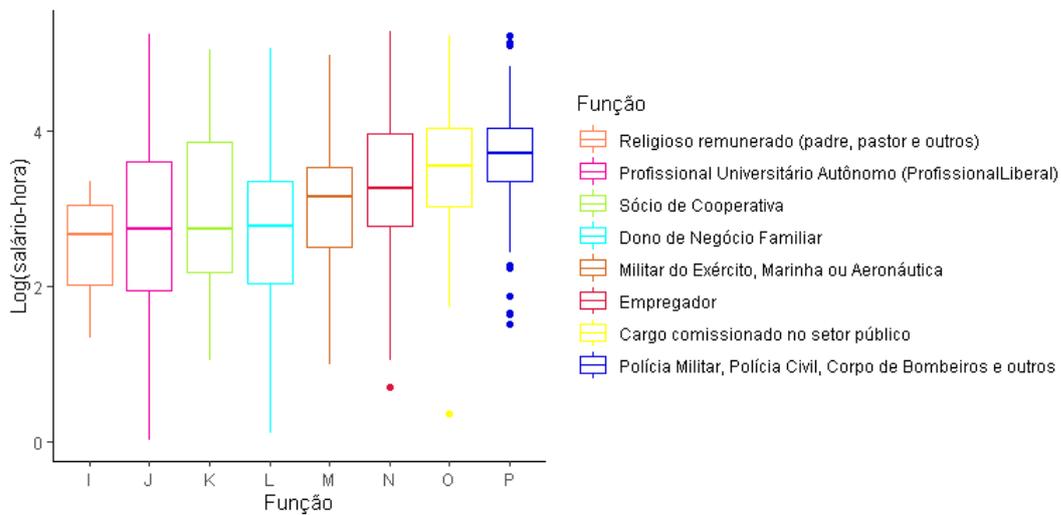


Gráfico 19: Distribuição de salários-hora, segundo função no trabalho, Distrito Federal - 2018



Ao comparar os Gráficos 18 e 19, nota-se uma diferença dos salários segundo cargo ou função desempenhada, principalmente porque as medianas e os quantis de 0,75 do Gráfico 19 excedem os do 18.

As funções mais bem pagas foram as de policiais e bombeiros, cargos comissionados no setor público, empregadores e militares. Ao passo que prestadores de serviço militar obrigatório, trabalhadores familiares e aprendizes foram as piores.

Outra característica que pode influenciar o salário é o setor de atividade da empresa. Nesse sentido, com o intuito de melhorar a visualização, já que ela também apresenta um número elevado de categorias, os dados foram divididos nos Gráficos 20 e 21, de modo que as áreas de trabalho aparecem em ordem crescente das medianas dos salários.

Gráfico 20: Distribuição de salários-hora, segundo setor de trabalho, Distrito Federal - 2018

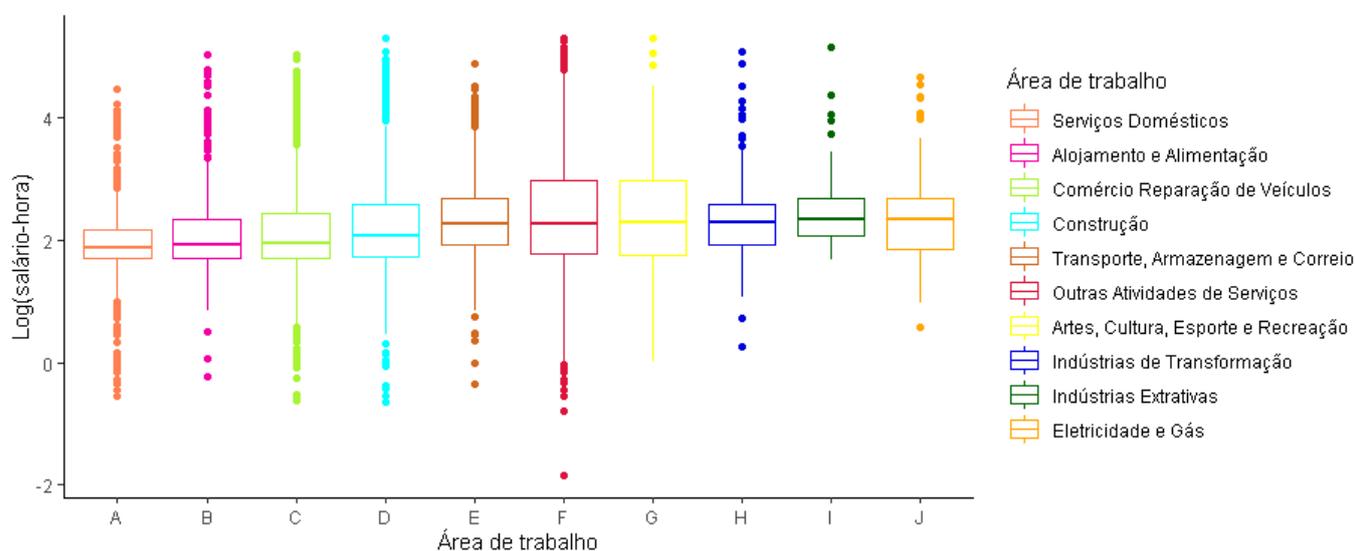
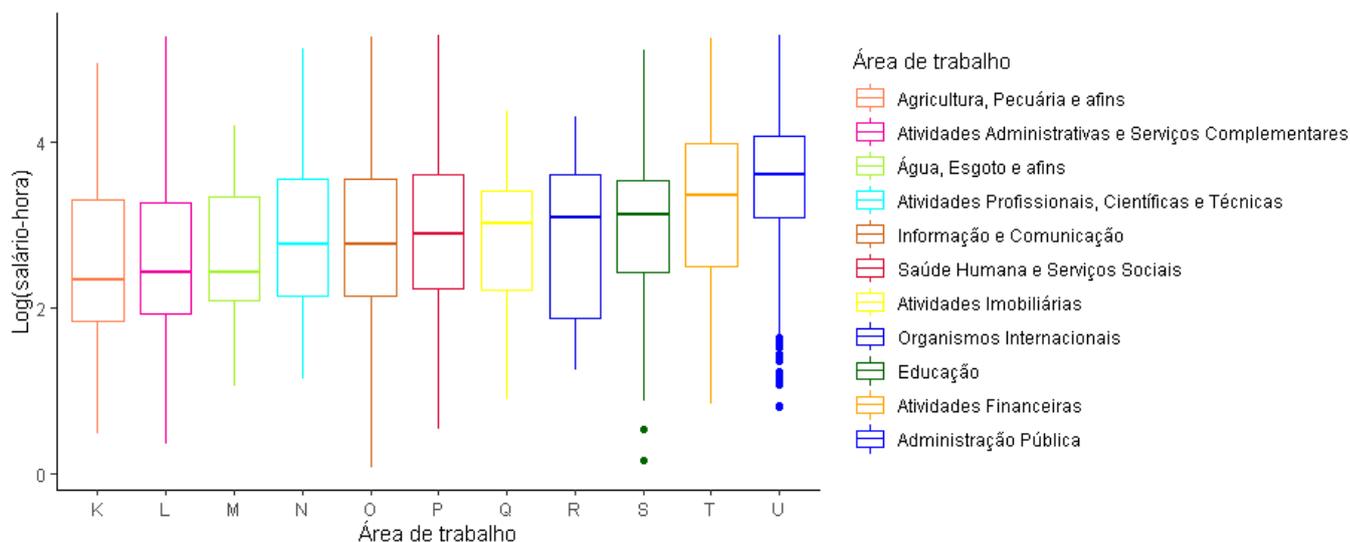


Gráfico 21: Distribuição de salários-hora, segundo setor de trabalho, Distrito Federal - 2018



Quanto ao setor de trabalho, segundo os Gráficos 20 e 21, também parece haver um contraste entre as rendas, da mesma maneira que a função no trabalho.

Aqui, os setores do Gráfico B apresentam maiores salários e, nesse caso, os setores mais bem remunerados são o de administração pública, de atividades financeiras e o de educação. Em contrapartida, aparecem o serviços domésticos, alojamento/alimentação e comércio de automotores com os menores salários.

Em se tratando de local de moradia, é possível verificar como o salário se distribui em função das regiões administrativas (RA's), uma vez que o Distrito Federal apresenta uma característica de concentração de renda em certos locais. Inclusive, sobre isso, a Tabela 8 mostra a classificação das RA's segundo a Pesquisa de Emprego e Desemprego (PED).

Tabela 7: Regiões administrativas segundo nível de renda, Distrito Federal - 2018

Nível de renda	Regiões administrativas
Alta	Jardim Botânico, Lago Norte, Lago Sul, Plano Piloto, Park Way e Sudoeste/Octogonal
Média-alta	Águas Claras, Candangolândia, Cruzeiro, Gama, Guará, Núcleo Bandeirante, Sobradinho, Sobradinho II, Taguatinga e Vicente Pires
Média-baixa	Brazlândia, Ceilândia, Planaltina, Riacho Fundo, Riacho Fundo II, SIA, Samambaia, Santa Maria e São Sebastião
Baixa	Fercal, Itapoã, Paranoá, Recanto das Emas, SCIA – Estrutural e Varjão

Fonte: Elaboração própria, dados da PED 2018 - DIEESE. ⁵

Para que seja possível identificar melhor o padrão de salários das 31 regiões admi-

⁵Disponível em: <https://www.dieese.org.br/analiseped/2018/201804pedbsb.html>

nistrativas (RA's), as informações de salário segundo RA foram distribuídas nos Gráficos 22, 23 e 24.

Gráfico 22: Distribuição de salários-hora, segundo RA de moradia, Distrito Federal - 2018 (A)

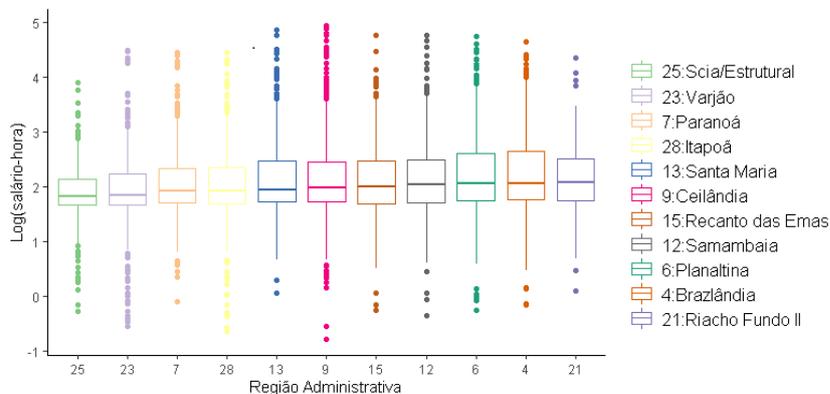
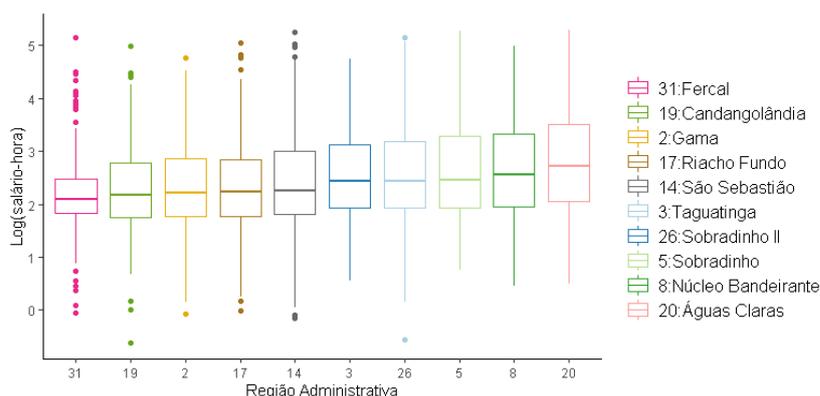


Gráfico 23: Distribuição de salários-hora, segundo RA de moradia, Distrito Federal - 2018 (B)



Dos Gráficos 22, 23 e 24, percebe-se que há discrepância do salário recebido em função do local de moradia, e também é interessante notar que as regiões tendem a aparecer conforme a classificação por renda da Pesquisa de Emprego e Desemprego (PED).

SCIA, Varjão, Paranoá, Itapoá, que são as regiões de baixa renda, possuem distribuição concentrada em valores baixos e aparecem no início do Gráfico. Já Sudoeste, Lago Sul, Jardim Botânico e Plano Piloto, de alta renda, aparecem no final e chegam a valores maiores que os outros locais.

Com relação à idade do trabalhador e ao tempo em que a pessoa está no emprego, são apresentados os Gráficos 25 e 26.

Analisando a figura 25, conclui-se que a idade pode ser relevante, já que a mediana dos salários aumenta conforme as categorias de idade.

Gráfico 24: Distribuição de salários-hora, segundo RA de moradia, Distrito Federal - 2018 (C)

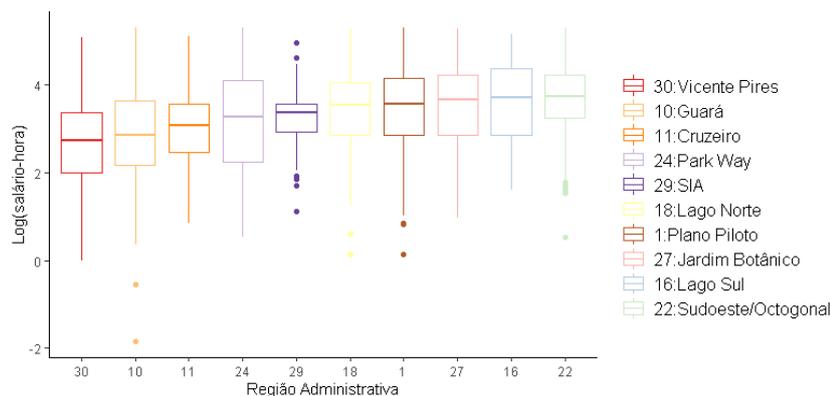
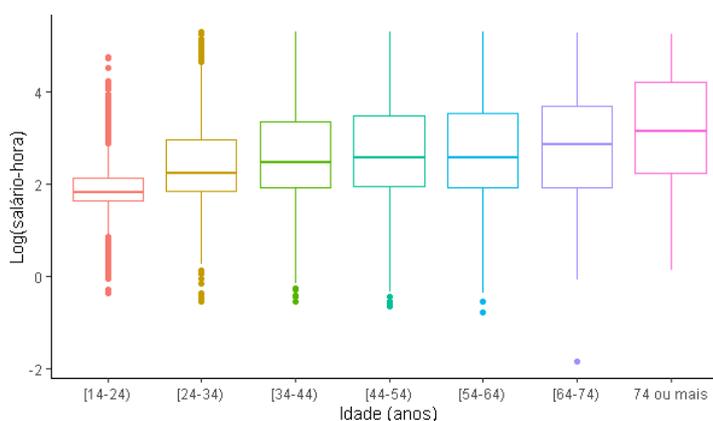


Gráfico 25: Distribuição de salários-hora, segundo idade, Distrito Federal - 2018



Isso pode estar relacionado à maior experiência no mercado de trabalho ou ao maior nível de escolaridade, já que há suspeita de que essas características tendem a aumentar o salário.

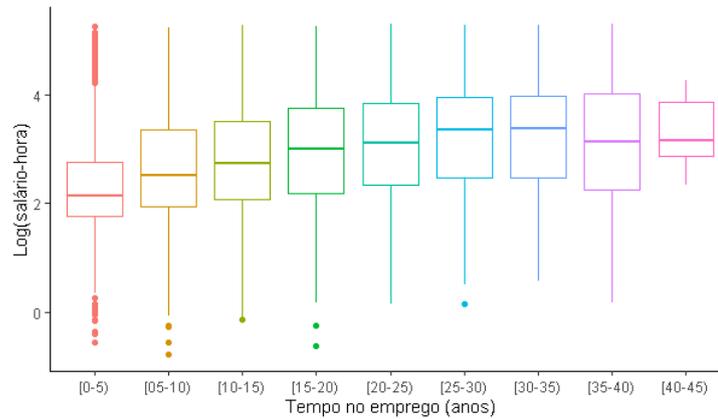
Observa-se que as menores remunerações são de pessoas com idades entre 14 e 23 anos, já que a haste superior do boxplot não chega ao valor 4 e as rendas atípicas, acima dela, chegam a valores menores que os das demais categorias.

Quanto ao tempo no emprego, conforme ilustrado na figura 26, verifica-se que os rendimentos aumentam até o período de 34 anos e decaem a partir de 35 anos. Além disso, as distribuições de salário para pessoas de 25 a 29 anos e de 30 a 34 anos são muito parecidas.

Também é interessante notar que a metade dos trabalhadores está há até 4 anos no emprego (Tabela 2) e a categoria “[0-5]” é a que apresenta os menores salários.

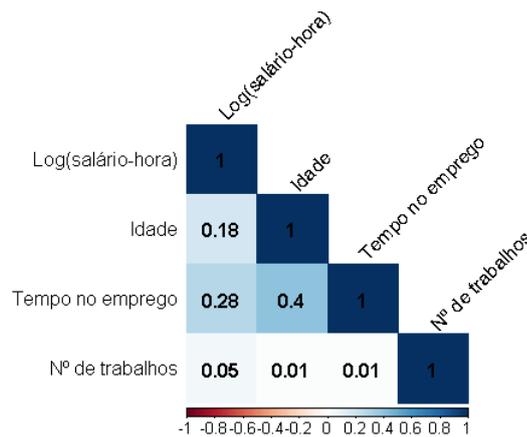
Por fim, é apresentada a matriz de correlação e resultados dos testes de correlação de Kendall entre as variáveis quantitativas estudadas, uma vez que elas não atendem as

Gráfico 26: Distribuição de salários-hora, segundo tempo no emprego, Distrito Federal - 2018



suposições para correlação de Pearson.

Gráfico 27: Matriz de Correlação de Kendall, Distrito Federal - 2018



O teste de correlação de Kendall estuda as seguintes hipóteses

$$\begin{cases} H_0 : \text{As variáveis são independentes } (\tau = 0). \\ H_1 : \text{As variáveis não são independentes } (\tau \neq 0). \end{cases}$$

Tabela 8: Resultados dos testes de Correlação de Kendall, Distrito Federal - 2018

Variáveis	Estatística	p.valor
Salário-hora x Idade	39,25	0
Salário-hora x Nº de trabalhos	59,13	0
Salário-hora x Tempo no emprego	8,71	0
Idade x Tempo no emprego	83,05	0
Idade x Nº de trabalhos	2,07	0,04
Tempo no emprego x Nº de trabalhos	2,18	0,03

Analisando a Tabela 8, conclui-se que, com nível de significância de 5%, existe evidência para rejeitar a hipótese de independência, assim considera-se que a correlação é significativa entre elas, uma vez que os p-valores são menores do que 0.05. Mas, consonante com o Gráfico 27, essas associações não são fortes, visto que $\tau = 0,4$, entre tempo no emprego e idade, é o maior valor obtido.

As relações foram estudadas nessa seção de forma exploratória, por meio de análises gráficas bivariadas. A regressão permite modelar as relações de forma multivariada, considerando o efeito de cada uma das variáveis explicativas no modelo considerando a presença das demais covariáveis. Esse será o objeto de estudo da próxima seção.

5.2 Análise de Regressão Linear

Nesta seção trataremos da regressão linear do salário em função das características dos trabalhadores. A regressão linear de uma variável Y em outras variáveis X_i , com $i = 1, \dots, p$, permite medir o quanto X_i influencia a média de Y , mantidas as demais X_{p-1} constantes, por esse motivo é de extrema relevância para estudos de tal finalidade, como é o caso do presente trabalho.

Vale ressaltar que por conta da assimetria da variável salário-hora, utilizaremos como variável resposta o logaritmo do salário-hora ($\log(\text{salário-hora})$) nas regressões linear e quantílica.

Além do mais, estamos interessados em modelar o salário-hora dos indivíduos (“Salário-hora”), no entanto só temos acesso ao salário dado que a pessoa trabalha (“Salário-hora” “Trabalhador”=Sim). Por isso, tendo como base (HECKMAN, 1979), incluiremos a variável inversa de Mills na regressão linear e na quantílica.

Para obter a inversa de Mills é necessário fazer a regressão Probit da variável “Trabalhador” em função de características que influenciam a entrada do indivíduo no mercado de trabalho. A regressão de trabalho é apresentada a seguir.

5.2.1 Regressão de trabalho - Probit

A regressão de trabalho leva em consideração apenas pessoas acima de 13 anos de idade, portanto o número de observações do estudo foi de 43.631. O resultado dessa análise é apresentado na Tabela 9.

Dada a Tabela 9, percebe-se que a maioria das variáveis abordadas influencia a

entrada ou não no mercado de trabalho, em 2018, no Distrito Federal, visto que a maioria dos pvalores estão abaixo de 5%.

Tabela 9: Resultados da regressão de trabalho, Distrito Federal - 2018

Variável	Estimativa
Intercepto	-3,0314***
Idade	0,1842***
Idade ²	-0,0022***
Estudantes	-0,2564***
Ensino Médio	0,1403***
Ensino Superior	0,3947***
Mulher	-0,2930***
Negro	0,0128
Deficiente	-0,4329***
Filho	0,1547***
Idoso	-0,3373***
Casado	0,0691***
Média-alta renda	0,0187
Média-baixa renda	0,0208
Baixa renda	0,1533***
Afazeres Domésticos	-0,0264***
Renda dom. s/ renda indivíduo(log)	0,0112***

Fonte: Elaboração própria, dados da PDAD 2018 - Codeplan.

Observações: *** indica significância a 0,1%;

** indica significância a 1% ;

*indica significância a 5% e

. indica significância a 10%.

Em relação ao sexo do indivíduo, nota-se que a probabilidade de mulheres estarem trabalhando é menor ⁷ que a dos homens, mantidas as demais covariáveis constantes.

Esse padrão de menor probabilidade de estar trabalhando também é constatado para aqueles que possuem alguma deficiência em relação aos que não têm e para os que moram com idosos em relação aos que não moram.

Já quem tem filho, reside na área de baixa renda, concluiu o ensino médio completo ou o ensino superior, possui maior probabilidade de estar inserido no mercado de

⁷O resultado é obtido a partir da diferença na probabilidade predita de estar trabalhando segundo a variação na variável de interesse, ou seja, computa-se $P(\text{Trabalha}=\text{Sim}|\mathbf{X}) = \Phi(\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n)$, para um valor de uma variável X e a mesma probabilidade para outro valor de X (em caso de variáveis qualitativas) ou para a variação de 1 unidade em X (em caso de variáveis quantitativas), mantidas as demais constantes, tendo os 2 resultados calcula-se a diferença entre eles. A obtenção dessa interpretação não é simples porque a relação entre o efeito obtido na probabilidade de Y dada uma diferença em X não é linear, mas esse procedimento pode ser obtido a partir da função `stats::predict()` do *software* R. Resultados negativos indicam que a categoria analisada possui probabilidade inferior a da categoria de referência, enquanto resultados positivos indicam probabilidade superior.

trabalho do que quem não tem essas características, provavelmente devido a maior dependência da remuneração (no caso da presença de filhos ou moradia na área de baixa renda) ou dos salários mais atrativos (no caso das pessoas com maior nível de escolaridade).

É interessante destacar que a probabilidade de alguém com ensino superior completo estar trabalhando supera a de alguém com escolaridade abaixo do ensino superior. Isso condiz com a crença de que um nível de instrução elevado facilita a inserção no mercado de trabalho e a oferta de salários mais atrativos incentiva esses trabalhadores a se manterem ativos.

Além disso, nota-se que a probabilidade de quem mora em regiões de baixa renda trabalhar supera a dos que vivem nas localidades de alta renda. Esse fato está atrelado, principalmente, às condições sociais a que esses públicos estão expostos, no que tange à qualidade de vida e sustento (básico) familiar, por exemplo, é mais comum ver jovens de menor renda trabalhando do que jovens de maior renda.

5.2.2 Seleção de variáveis

Para a regressão de salários não foram levados em conta os seguintes registros do banco de dados: pessoas que possuíam menos de 14 anos (12.144 observações); que trabalham e apresentam renda 0 (Tabela 10) e que não trabalham (29.985 observações) ou não responderam. No total foram usadas 18.922 observações nessa análise.

Tabela 10: Pessoas que trabalham e não têm renda, segundo função no trabalho, Distrito Federal - 2018

Setor de trabalho	Frequência
Aprendiz	11
Conta Própria ou Autônomo	85
Dono de Negócio Familiar	5
Empregado (Exceto Empregado Doméstico)	91
Empregado Doméstico	13
Empregado Temporário	2
Empregador	3
Estágio Remunerado	16
Polícia Militar, Polícia Civil, Corpo de Bombeiros e outros	1
Profissional Universitário Autônomo (Profissional Liberal)	1
Religioso remunerado (padre, pastor e outros)	2
Trabalhador familiar sem remuneração salarial	11
NA	3
Total	244

Fonte: Elaboração própria, dados da PDAD 2018 - Codeplan.

Além disso, foram analisados 10 modelos, incluindo o proposto por Rosa, Quirino e Xavier (2021), e para a seleção de variáveis levou-se em consideração o critério de Akaike (AIC), o critério Bayesiano (BIC) e R^2 . Os modelos ajustados estão apresentados no Apêndice A e a comparação entre eles está apresentada na Tabela 11.

O modelo escolhido foi o de número 7, que aborda as variáveis da Tabela 13. Conforme a Tabela 11, os modelos 2 a 5 apresentaram menores AIC e BIC, porém abordam a variável “Função no trabalho” que apresentou muitas características não significativas, ao nível de significância de 5%, o que leva à consideração de analisar o modelo sem ela. O modelo 5 também leva em consideração “Setor da economia”, mas ela não foi significativa a 5%, por isso, optamos por retirá-la do modelo.

As variáveis “Setor de trabalho” e “Função no trabalho” foram investigadas em conjunto nos modelos 3 e 4, e em separado nos modelos de 6 a 9.

Comparando os modelos 6 e 3, não há um ganho significativo da inserção de “Função de trabalho” no modelo, visto que o aumento em R^2 é ínfimo. O mesmo ocorre quando comparamos 4 e 7. Em relação ao acréscimo de “Setor de trabalho” (3 e 4) ao modelo com “Função no trabalho” (8 e 9), também não há ganhos expressivos, mas essa parece ser uma variável relevante, já que na simulação de análise de variância (ANOVA - Tabela 12) para comparação do modelo com “Setor de trabalho” (7) e o sem ambas as variáveis (10), rejeita-se a hipótese nula com $\alpha = 5\%$. Portanto, o modelo sem as 2 variáveis não é escolhido.

Tabela 11: Critérios de Akaike, Bayesiano e R^2 para seleção de variáveis segundo modelo usado, Distrito Federal - 2018

Modelo	AIC	BIC	R^2	R^2 corrigido
Modelo 1	37053,89	37171,64	0,5384	0,5384
Modelo 2	34297,23	34720,59	0,5917	0,5906
Modelo 3	34300,25	34676,86	0,5914	0,5904
Modelo 4	34458,29	34740,53	0,5874	0,5867
Modelo 5	34864,01	35052,17	0,5779	0,5774
Modelo 6	35085,18	35367,71	0,5829	0,5822
Modelo 7	35086,29	35345,27	0,5828	0,5821
Modelo 8	34800,61	35043,7	0,5806	0,58
Modelo 9	34920,63	35093,15	0,5775	0,5771
Modelo 10	35566,86	35716,02	0,5734	0,573

Fonte: Elaboração própria, dados da PDAD 2018 - Codeplan.

Observação: As descrições dos modelos se encontram na seção A do apêndice.

Tabela 12: ANOVA para seleção de modelos, Distrito Federal - 2018

Modelo	GL	SSE	MSE	Estatística F	P-valor
Sem setor e sem função de trabalho	18947	7229,1	3,007	8,056	0
Com setor de trabalho	18887	7048,7			

Fonte: Elaboração própria, dados da PDAD 2018 - Codeplan.

O resultado da regressão linear de salários para o Distrito Federal em 2018 é apresentado na Tabela 13.

Tabela 13: Resultados da regressão linear de log(salários-hora), Distrito Federal - 2018

Variáveis	Estimativa
Intercepto	1,9782***
Idade	0,0230***
Idade ²	-0,0002***
Tempo no emprego	0,0142***
2 trabalhos ou mais	0,1403***
Mulher	-0,1473***
Negro	-0,0776***
Deficiente	-0,0338
Estudante	0,1473***
Ensino médio	0,1172***
Superior	0,7185***
Trabalha no setor público	0,4335***
Trabalha no setor informal	-0,0971***
Informação e Comunicação	0,2486***
Atividades Financeiras	0,3895***
Atividades Imobiliárias	0,3429***
Atividades Profissionais, Científicas e Técnicas	0,3170***
Atividades Administrativas e Serviços Complementares	0,1349***
Administração Pública	0,3421***
Educação	0,1600***
Saúde Humana e Serviços Sociais	0,3216***
Indústrias de Transformação	0,2265***
Eletricidade e Gás	0,1568 *
Comércio Reparação de Veículos	0,0680 **
Transporte, Armazenagem e Correio	0,1265***
Alojamento e Alimentação	0,0725 *
Outros setores	0,0988***
Média-alta renda	-0,3980***
Média-baixa renda	-0,5783***
Baixa renda	-0,6079***
Trabalha no Plano Piloto	0,0813***
IMR	-0,1364***

Fonte: Elaboração própria, dados da PDAD 2018 - Codeplan.

Observações: *** indica significância a 0,1%;

** indica significância a 1% ;

*indica significância a 5% e

. indica significância a 10%.

A partir da Tabela 13 observamos que as mulheres ganham 86,3%⁸ do que os homens recebem, ou seja, recebem 13,7%⁹ a menos. Já negros têm salários 7,5% menores que não negros.

Graus de instrução completos também garantem um salário melhor, uma vez que quem tem ensino médio completo recebe 12,4% a mais que os que não têm e os que possuem ensino superior, 105,1%.

Também não há diferença significativa entre os salários médios de pessoas com e sem deficiência.

A cada um ano a mais de idade, a média do salário-hora sobe 2,33% e a cada um ano a mais no trabalho esse aumento é de 1,43%.

Quem tem mais de um trabalho garante um salário médio 15% maior do que quem não tem. O salário médio dos trabalhadores informais é 9,3% menor que o dos formais, enquanto o dos que trabalham no setor público é 54,3% maior.

Conforme esperado, o salário apresenta diferenças em relação ao local de moradia das pessoas, sendo que aqueles que residem em regiões de alta renda apresentam salários significativamente maiores que os que residem em localidades com níveis de renda inferiores.

No que se refere ao setor de trabalho, nota-se que todas as áreas apresentam salários estatisticamente diferentes da de serviços domésticos (categoria de referência), de forma que todas as outras possuem salários superiores e as que mais se afastam são os setores financeiro (47,6% a mais), imobiliário (40,9% a mais) e de administração pública (40,8% a mais).

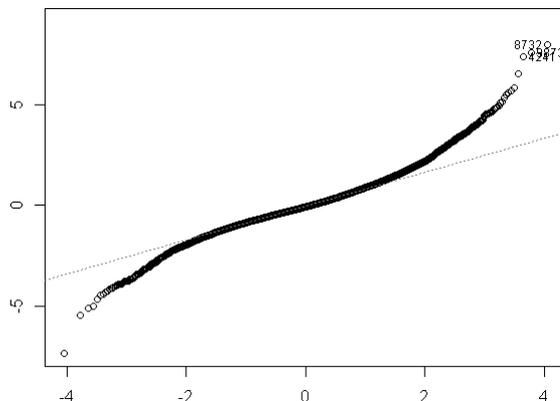
O coeficiente para inversa de Mills é significativo a 5%, indicando que considerar a questão ocupacional na análise de salários é importante, pois trata o viés de seleção amostral decorrente da influência das características pessoais na inserção dos indivíduos no mercado de trabalho. Além do mais, o fato de esse coeficiente ser negativo, indica que a ausência da inversa de Mills na equação salarial geraria vieses para baixo, ou seja, as estimativas obtidas da regressão seriam menores do que realmente são, isso acarretaria em erros na quantificação das diferenças salariais em função das características analisadas.

O Gráfico 28 traz um diagnóstico para a normalidade dos resíduos da regressão aplicada. Diante dele, é possível perceber que, mesmo com a transformação logarítmica

⁸A regressão foi aplicada em $\log(\text{salário-hora})$, portanto para que possamos interpretar em função do salário-hora precisamos aplicar exponencial às estimativas obtidas, nesse caso fazemos $\exp(-0,1473)$.

⁹Resulta da comparação de salários e é obtido por $\exp(-0,1472)-1$.

Gráfico 28: Comparação entre quantis obtidos para os resíduos e os de uma Distribuição Normal, Distrito Federal - 2018



da variável salário-hora, os resíduos desviam da suposição de normalidade. Além disso, parecem apresentar uma distribuição de caudas mais pesadas, como uma T-Student, uma vez que existe um afastamento pronunciado entre os pontos (resíduos) e a reta teórica nos intervalos $[-4, -2,5]$ e $[2,5, 4]$. Esse resultado já era esperado dada a assimetria da variável $\log(\text{salário-hora})$ (Gráfico 10).

Ainda nesse sentido, foi feita uma simulação por bootstrap do teste de Shapiro-Wilk, cuja hipótese nula, nesse caso, é de normalidade dos resíduos. Para tanto foram retiradas 10.000 amostras (com 5.000 observações cada) da amostra original de resíduos para o modelo obtido (Tabela 13). Em cada uma das amostras foi aplicado o teste de Shapiro-Wilk, que, considerando nível de significância de 5%, rejeitou a hipótese nula em todas elas.

A regressão linear é construída com base em pressupostos, como mencionado acima, e se algum deles não é aceito, podem existir problemas na estimação dos coeficientes do modelo. Esse fato reitera a importância da aplicação de uma regressão quantílica, já que esse método de análise não requer suposições sobre a distribuição da variável resposta.

5.3 Análise de Regressão Quantílica

Por fim, abordaremos neste tópico a regressão quantílica, que permite a extensão da análise ao longo da distribuição de salários, para além da média, propiciando, assim, uma visão mais ampla das relações existentes em diferentes níveis de renda.

Além disso, a regressão em quantis é uma boa alternativa quando há presença de valores discrepantes (outliers) ou não atendimento dos pressupostos de normalidade dos resíduos e equivariância.

O resultado da regressão quantílica de salários é apresentado na Tabela 14. Foram utilizadas as mesmas covariáveis selecionadas no modelo de regressão linear para a média (ou seja, modelo 7) e ajustados os quantis $\tau = 0, 1$, $\tau = 0, 25$, $\tau = 0, 5$, $\tau = 0, 75$ e $\tau = 0, 9$.

Tabela 14: Resultados da regressão quantílica de log(salários-hora), Distrito Federal - 2018

Variáveis	MOQ	Q=0,1	Q=0,25	Q=0,5	Q=0,75	Q=0,9
Intercepto	1,9782***	1,7119***	1,8694***	1,8958***	1,8930***	2,2146***
Idade	0,0230***	0,0129 **	0,0151***	0,0241***	0,0393***	0,0385***
Idade ²	-0,0002***	-0,0001 .	-0,0001 **	-0,0002***	-0,0004***	-0,0003***
Tempo no emprego	0,0142***	0,0102***	0,0129***	0,0159***	0,0169***	0,0171***
2 trabalhos ou mais	0,1403***	0,0700 *	0,0766 *	0,1469***	0,1193	0,2281 **
Mulher	-0,1473***	-0,0742***	-0,0935***	-0,1379***	-0,2057***	-0,1734***
Negro	-0,0776***	-0,0613***	-0,0622***	-0,0559***	-0,0727***	-0,0875***
Deficiente	-0,0338	-0,0580 .	-0,0218	-0,0251	-0,0191	-0,0654 **
Estudante	0,1473***	0,0853***	0,1359***	0,1427***	0,1666***	0,1580***
Ensino médio	0,1172***	0,0769***	0,0958***	0,1150***	0,1189***	0,0971***
Superior	0,7185***	0,4513***	0,5863***	0,7552***	0,8517***	0,8245***
Trabalha no setor público	0,4335***	0,6069***	0,5866***	0,5007***	0,3896***	0,3109***
Trabalha no setor informal	-0,0971***	-0,2786***	-0,1150***	-0,0572***	0,0043	0,0665***
Informação e Comunicação	0,2486***	0,1400***	0,1744***	0,2135***	0,2695***	0,3987***
Atividades Financeiras	0,3895***	0,1970***	0,2731***	0,3647***	0,4238***	0,5383***
Atividades Imobiliárias	0,3429***	0,1673	0,2712***	0,2862 **	0,4261 *	0,6293***
At. Profissionais, Cient. e Técnicas	0,3170***	0,2121***	0,2173***	0,2805***	0,3021***	0,4643***
At. Administ. e Serv. Complementares	0,1349***	0,0811 **	0,0971***	0,1014***	0,1658***	0,2647***
Administração Pública	0,3421***	0,1785***	0,2715***	0,3022***	0,3938***	0,4812***
Educação	0,1600***	0,1226***	0,1255***	0,1126***	0,1512***	0,1683***
Saúde Humana e Serviços Sociais	0,3216***	0,1710***	0,2443***	0,2563***	0,3355***	0,4555***
Indústrias De Transformação	0,2265***	0,1539 *	0,1626***	0,1652***	0,3057***	0,3580 *
Eletricidade e Gás	0,1568 *	0,0182	0,1936 *	0,1502***	0,1384***	0,1992
Comércio Reparação De Veículos	0,0680 **	0,0150	0,0241	0,0290 .	0,0643 **	0,1573***
Transporte, Armazenagem e Correio	0,1265***	0,0518	0,0751 **	0,1443***	0,1733***	0,1967***
Alojamento e Alimentação	0,0725 *	-0,0067	0,0240	0,0086	0,0323	0,1630 .
Outros setores	0,0988***	0,0154	0,0312 *	0,0503 **	0,1156***	0,2420***
Média-alta renda	-0,3980***	-0,2735***	-0,3775***	-0,3982***	-0,4403***	-0,4551***
Média-baixa renda	-0,5783***	-0,3826***	-0,4903***	-0,5534***	-0,6461***	-0,6994***
Baixa renda	-0,6079***	-0,4063***	-0,4957***	-0,5747***	-0,6722***	-0,7478***
Trabalha no Plano Piloto	0,0813***	0,0642***	0,0545***	0,0753***	0,0667***	0,0475 **
IMR	-0,1364***	-0,2296***	-0,1661***	-0,1072***	-0,0443	-0,0461

Fonte: Elaboração própria, dados da PDAD 2018 - Codeplan.

Observações: *** indica significância a 0,1%;

** indica significância a 1% ;

*indica significância a 5% e

. indica significância a 10%.

Observando a Tabela 14, percebe-se que, em 2018, as características de idade, de tempo no emprego, de ser mulher, de ser negro, estar estudante, ter ensino médio/ensino superior completos, trabalhar no setor público, no setor informal e no Plano Piloto, bem como o local de moradia acarretam em diferenças salariais em relação aos que não possuem

tais características em todos os níveis salariais, ou seja, desde os menores salários até os maiores.

No que diz respeito ao sexo do indivíduo, conforme a Tabela 14 e o Gráfico 29, as mulheres recebem menos que os homens em todos os quantis de salário.

Além disso, essa diferença se acentua com o aumento da faixa salarial, já que para salários: do extremo inferior da distribuição (quantil 0,1), elas recebem 7,2%¹⁰ a menos; do quantil 0,5, 12,9% menos e do extremo superior (quantil 0,9), 15,9% menos. É interessante destacar que a maior diferença percebida entre sexos é de 18,6% para o quantil 0,75 de salários, que representa os salários acima da média, mas não tão altos.

Agora, em relação à raça, negros recebem menos que não negros em todos os quantis de renda (Tabela 14 e Gráfico 30), porém em proporções menores que o sexo. Observa-se um padrão um pouco diferente, sendo que a menor discrepância é de -5,4% para a mediana dos salários e a maior de -8,38% para a cauda superior da distribuição de salários (quantil 0,9).

Existe uma diferença significativa entre a receita de pessoas com e sem alguma deficiência apenas no quantil 0,9, que representa a cauda superior da distribuição de salários. Nesse quantil pessoas com deficiência recebem salário 6,33% menor que pessoas sem.

A escolaridade em 2018 é um fator que influencia no salário recebido, já que quem estuda, tem ensino médio completo e quem tem ensino superior completo possui ganhos mais elevados que aqueles que não dispõem dessas características. É surpreendente como ensino superior completo aumenta o salário em todos os níveis de renda, mas principalmente para os quantis 0,75 e 0,90, com salários 134,4% e 128,1% maiores, respectivamente (Gráfico 31).

Os que trabalham no setor público também apresentam maiores salários, contudo a relação é decrescente. Assim, para a parte inferior da distribuição de rendas, quem está nesse setor de trabalho tem uma remuneração 83,5% maior que quem não está. Enquanto para a parte superior, essa diferença é de 36,5% (Tabela 14).

Para regime de trabalho informal (Gráfico 32), observa-se uma inversão nos salários conforme o nível de renda. Os trabalhadores informais com rendas pertencentes à cauda inferior da distribuição de salários (quantil 0,1 de salários) chegam a ganhar 24,3% a menos que os formais. Já aqueles com rendas pertencentes à mediana dos salários ganham 5,6% menos. E aqueles com rendas pertencentes ao quantil 0,9 (cauda superior)

¹⁰Resulta da comparação de salários e é obtido por $\exp(-0,0742)-1$.

recebem 6,9% a mais que os trabalhadores formais. Essa inversão pode estar vinculada ao setor de trabalho escolhido/disponível para esses trabalhadores em função da renda ou grau de influência que possuem.

Trabalhar no Plano Piloto também promove o salário, principalmente, para aqueles que apresentam rendas correspondentes à mediana da distribuição dos salários (quantil 0,5). Nesse quantil, quem trabalha no Plano Piloto recebe 7,8% a mais do que quem não trabalha no centro da cidade.

A quantidade de trabalho também influencia na recompensa trabalhista, como era de se esperar. E constata-se que em todos os níveis de renda, ter dois trabalhos ou mais gera um ganho superior a 7% em relação aos que têm apenas 1 trabalho. Inclusive, esse ganho cresce conforme a faixa salarial aumenta, chegando a 25,6% no quantil 0,9.

Tanto a idade quanto o tempo no emprego geram retornos positivos e que crescem conforme o aumento da faixa salarial. De forma que cada um ano a mais de idade possibilita um aumento entre 1,3% e 4% no salário-hora. Enquanto cada ano a mais no emprego aumenta o salário-hora em no mínimo 1,02% e no máximo 1,73%.

No tocante à área de trabalho, serviços domésticos geram salário pouco maior (de 0,7%) apenas se comparados com alojamento e alimentação no menor quantil de salários (quantil 0,1). Todos os outros setores têm melhor remuneração e essa discrepância se acentua conforme o nível de salário cresce. O setor imobiliário, financeiro e de administração pública continuam sendo os melhores pagos em relação aos serviços domésticos. Essa diferença é grande e varia de: 18,21 a 87,62%; 21,78 a 71,31% e 19,55 a 61,81%, nessa ordem.

Comparativamente, em relação às pessoas que residem em região de alta renda (Grupo 1 da PED), aquelas que moram em regiões de média-alta, média-baixa e baixa, têm salários inferiores. É interessante destacar que essa diferença se agrava com a diferença nos níveis de renda e no poder aquisitivo entre as regiões. A exemplo disso, para a cauda inferior da distribuição de salários (quantil 0,1): locais de baixa renda apresentam salários 33,4% menores; de média-baixa, 31,8% menores e de média-alta, 23,9% menores. Já para a cauda superior (quantil 0,9) essa diferença é de 52,66%, 50,31% e 36,56%, respectivamente.

Gráfico 29: Coeficiente da regressão quantílica de log(salário-hora) para sexo, Distrito Federal - 2018

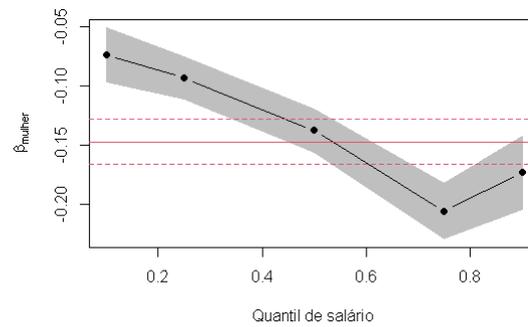


Gráfico 30: Coeficiente da regressão quantílica de log(salário-hora) para raça, Distrito Federal - 2018

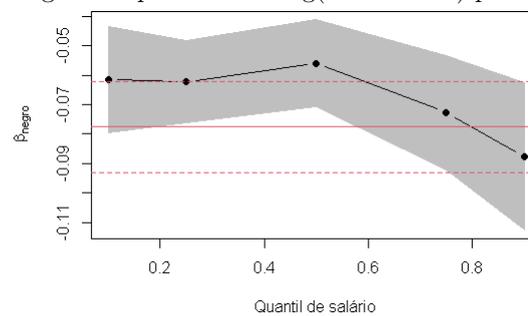


Gráfico 31: Coeficiente da regressão quantílica de log(salário-hora) para ensino superior, Distrito Federal - 2018

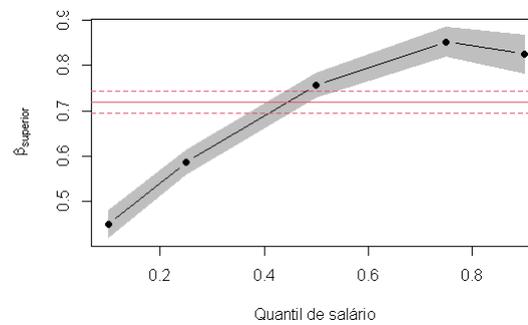
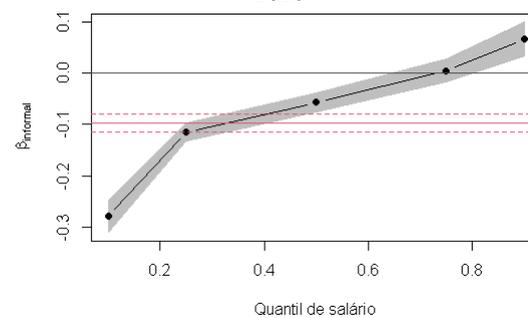


Gráfico 32: Coeficiente da regressão quantílica de log(salário-hora) para setor informal, Distrito Federal - 2018



6 Conclusão

Dado o exposto, primeiro, percebemos que a regressão quantílica é relevante de ser abordada no contexto deste estudo, uma vez que seus resultados mostraram que os efeitos de características individuais no salário recebido eram desiguais nos diferentes níveis de renda.

Alguns resultados interessantes são: existe uma certa concordância entre os resultados obtidos neste trabalho e no de Rosa, Quirino e Xavier (2021), como é o caso da questão de sexo do indivíduo, mas também há uma inversão em alguns deles, como para trabalhadores informais.

Quanto ao sexo, tanto neste ensaio quanto no outro, obteve-se que remunerações femininas são inferiores às masculinas, quando trata-se da média de salários essa diferença é de 14,2% (ROSA; QUIRINO; XAVIER, 2021), já na mediana é de 12,9% e, no quantil 0,75, chega a 18,6%.

A média salarial para negros é 13% menor que a de não negros (ROSA; QUIRINO; XAVIER, 2021), quando analisa-se nos quantis de renda, as diferenças existem nesse mesmo sentido, mas em menor expressividade que esse valor.

Tanto na remuneração média quanto nos quantis de renda, os retornos à escolaridade são bem diferentes dependendo do nível de instrução. Para quem tem ensino médio completo, o retorno médio é 20%¹¹ maior e para superior é 181%¹² maior, o quantil 0,75 apresenta as maiores diferenças para ambos os níveis, médio (retorno 12,6% maior) e superior (retorno 134,4% maior).

Um ano a mais no emprego leva a um aumento em 1,6%¹³ no salário-hora médio. Um valor próximo a esse é observado no quantil 0,9 de salários (1,73%) e é no mínimo de 1,02% no quantil 0,1.

Aqueles que pertencem ao setor informal mostram um ponto de vista interessante, visto que o esperado é que essas pessoas tenham menores retornos que os presentes no formal. Quando se olha para a média e nos quantis inferiores da distribuição de salário-hora, é exatamente esse resultado que se obtém, mais precisamente, retorno médio 16% menor¹⁴. Contudo no quantil 0,9 de salários-hora, trabalhadores informais recebem 6,9% mais que os formais.

¹¹Cálculo de $100 * (\exp(0,185) - 1) = 20,3$: Tabela 12 (ROSA; QUIRINO; XAVIER, 2021).

¹²Cálculo de $100 * (\exp(1,035) - 1) = 181,5$: Tabela 12 (ROSA; QUIRINO; XAVIER, 2021).

¹³Cálculo de $100 * (\exp(0,016) - 1) = 1,6$: Tabela 12 (ROSA; QUIRINO; XAVIER, 2021).

¹⁴Cálculo de $100 * (\exp(-0,178) - 1) = -16,3$: Tabela 12 (ROSA; QUIRINO; XAVIER, 2021).

Por fim, verifica-se que a área de trabalho, a quantidade de empregos e o local de residência (segundo poder aquisitivo) influenciam a recompensa angariada, da forma que se espera, ou seja, verificam-se maiores salários em regiões de alta renda e entre aqueles que têm mais de um trabalho, porém setor de serviço doméstico ainda é desvalorizado em relação aos demais.

Referências

- HECKMAN, J. J. Sample Selection Bias as a Specification Error. *The Econometric Society*, v. 47, n. 1, p. 153–161, 1979.
- KOENKER, R. *Quantile regression*. [S.l.]: Cambridge University Press, 2005.
- KOENKER, R.; BASSETT, G. Regression quantiles. *Econometrica*, v. 46, p. 33–50, 1978.
- KUTNER, M. H. et al. *Applied linear statistical models*. [S.l.]: McGraw-Hill New York, 2005.
- RIBEIRO, B. C.; KOMATSU, B. K.; FILHO, N. M. Diferenciais salariais por raça e gênero para formados em escolas públicas ou privadas. *Centro de Gestão e Políticas Públicas*, 2020.
- ROSA, T. M.; QUIRINO, I. S. R.; XAVIER, T. A. Salário feminino no distrito federal segundo a pdad 2018. *Companhia de Planejamento do Distrito Federal*, 2021.
- SANTOS, B. R. d. Modelos de regressão quantílica. *Dissertação (Mestrado em Estatística) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo*, 2012.
- SOUSA, L. P. d.; GUEDES; ROCHA, D. A desigual divisão sexual do trabalho: um olhar sobre a última década. *Estudos Avançados [online]*, v. 30, p. 123–139, 2016.

Apêndice

A Seleção de variáveis - Regressão linear

Modelo 1 - Variáveis apresentadas no trabalho de Rosa, Quirino e Xavier (2021):

$$\text{Log(salário-hora)} = \text{Idade} + \text{Idade}^2 + \text{Tempo no emprego} + \text{Mulher} + \text{Negro} + \text{Deficiente} + \text{Estudante} + \text{Ensino médio} + \text{Superior} + \text{Setor público} + \text{Setor informal} + \text{Trabalha no Plano Piloto} + \text{Inversa de Mills}.$$

Modelo 2 - Inclusão das variáveis “Qtd. trabalho”, “Grupos de RA”, “Setor de trabalho” e “Função de trabalho” com todas as categorias:

$$\text{Log(salário-hora)} = \text{Idade} + \text{Idade}^2 + \text{Tempo no emprego} + \text{Qtd. trabalho} + \text{Mulher} + \text{Negro} + \text{Deficiente} + \text{Estudante} + \text{Ensino médio} + \text{Superior} + \text{Setor público} + \text{Setor informal} + \text{Setor de trabalho} + \text{Função de trabalho} + \text{Grupos de RA} + \text{Trabalha no Plano Piloto} + \text{Inversa de Mills}.$$

Obs: Muitas categorias não significativas para “Função de trabalho”, considerando nível de significância de 5%.

Modelo 3 - Variáveis “Setor de trabalho” e “Função de trabalho” com categorias contendo menos de 100 observações agrupadas em “outros”:

$$\text{Log(salário-hora)} = \text{Idade} + \text{Idade}^2 + \text{Tempo no emprego} + \text{Qtd. trabalho} + \text{Mulher} + \text{Negro} + \text{Deficiente} + \text{Estudante} + \text{Ensino médio} + \text{Superior} + \text{Setor público} + \text{Setor informal} + \text{Setor de trabalho A} + \text{Função de trabalho A} + \text{Grupos de RA} + \text{Trabalha no Plano Piloto} + \text{Inversa de Mills}.$$

Modelo 4 - Variáveis “Setor de trabalho” e “Função de trabalho” com categorias contendo menos de 100 observações ou que não foram significativas no modelo 2 agrupadas em “outros”:

$$\text{Log(salário-hora)} = \text{Idade} + \text{Idade}^2 + \text{Tempo no emprego} + \text{Qtd. trabalho} + \text{Mulher} + \text{Negro} + \text{Deficiente} + \text{Estudante} + \text{Ensino médio} + \text{Superior} + \text{Setor público} + \text{Setor informal} + \text{Setor de trabalho B} + \text{Função de trabalho B} + \text{Grupos de RA} + \text{Trabalha no Plano Piloto} + \text{Inversa de Mills}.$$

Modelo 5 - Variável “Setor da economia” e “Função de trabalho” com categorias contendo menos de 100 observações ou que não foram significativas no modelo 1 agrupadas em “outros”:

$$\text{Log}(\text{salário-hora}) = \text{Idade} + \text{Idade}^2 + \text{Tempo no emprego} + \text{Qtd. trabalho} +$$
$$\text{Mulher} + \text{Negro} + \text{Deficiente} + \text{Estudante} + \text{Ensino médio} + \text{Superior} +$$
$$\text{Setor público} + \text{Setor informal} + \text{Setor da economia} + \text{Função de trabalho B} +$$
$$\text{Grupos de RA} + \text{Trabalha no Plano Piloto} + \text{Inversa de Mills.}$$

Obs: Variável “Setor da economia” não foi significativa”, considerando nível de significância de 5%.

Modelo 6 - Variável “Setor de trabalho” com categorias contendo menos de 100 observações agrupadas em “outros”:

$$\text{Log}(\text{salário-hora}) = \text{Idade} + \text{Idade}^2 + \text{Tempo no emprego} + \text{Qtd. trabalho} +$$
$$\text{Mulher} + \text{Negro} + \text{Deficiente} + \text{Estudante} + \text{Ensino médio} + \text{Superior} +$$
$$\text{Setor público} + \text{Setor informal} + \text{Setor de trabalho} + \text{Grupos de RA} +$$
$$\text{Trabalha no Plano Piloto} + \text{Inversa de Mills.}$$

Modelo 7 - Variável “Setor de trabalho” com categorias contendo menos de 100 observações ou que não foram significativas no modelo 2 agrupadas em “outros”:

$$\text{Log}(\text{salário-hora}) = \text{Idade} + \text{Idade}^2 + \text{Tempo no emprego} + \text{Qtd. trabalho} +$$
$$\text{Mulher} + \text{Negro} + \text{Deficiente} + \text{Estudante} + \text{Ensino médio} + \text{Superior} +$$
$$\text{Setor público} + \text{Setor informal} + \text{Setor de trabalho B} + \text{Grupos de RA} +$$
$$\text{Trabalha no Plano Piloto} + \text{Inversa de Mills.}$$

Modelo 8 - Variável “Função de trabalho” com categorias contendo menos de 100 observações agrupadas em “outros”:

$$\text{Log}(\text{salário-hora}) = \text{Idade} + \text{Idade}^2 + \text{Tempo no emprego} + \text{Qtd. trabalho} +$$
$$\text{Mulher} + \text{Negro} + \text{Deficiente} + \text{Estudante} + \text{Ensino médio} + \text{Superior} +$$
$$\text{Setor público} + \text{Setor informal} + \text{Função de trabalho} + \text{Grupos de RA} +$$
$$\text{Trabalha no Plano Piloto} + \text{Inversa de Mills.}$$

Modelo 9 - Variável “Função de trabalho” com categorias contendo menos de 100 observações ou que não foram significativas no modelo 2 agrupadas

em “outros”:

$$\text{Log}(\text{salário-hora}) = \text{Idade} + \text{Idade}^2 + \text{Tempo no emprego} + \text{Qtd. trabalho} +$$

$$\text{Mulher} + \text{Negro} + \text{Deficiente} + \text{Estudante} + \text{Ensino médio} + \text{Superior} +$$

$$\text{Setor público} + \text{Setor informal} + \text{Função de trabalho B} + \text{Grupos de RA} +$$

$$\text{Trabalha no Plano Piloto} + \text{Inversa de Mills.}$$

Modelo 10 - Sem as variáveis “Setor de trabalho” e “Função de trabalho”:

$$\text{Log}(\text{salário-hora}) = \text{Idade} + \text{Idade}^2 + \text{Tempo no emprego} + \text{Qtd. trabalho} +$$

$$\text{Mulher} + \text{Negro} + \text{Deficiente} + \text{Estudante} + \text{Ensino médio} + \text{Superior} +$$

$$\text{Setor público} + \text{Setor informal} + \text{Grupos de RA} +$$

$$\text{Trabalha no Plano Piloto} + \text{Inversa de Mills.}$$

B Gráficos: Regressão quantílica x Regressão linear

A seguir são apresentados os gráficos com resultados das regressões quantílicas e da regressão linear para $\log(\text{salário-hora})$. A partir deles é possível ter um melhor detalhamento das diferenças captadas por cada uma das técnicas. Como é perceptível, algumas das variáveis analisadas apresentam diferentes efeitos para os diferentes níveis de renda, que não seriam captados pela regressão linear, isso reitera a importância da regressão quantílica nesse contexto.

Gráfico 33: Estimativas e intervalos de confiança da regressão quantílica e da linear para Log(salário-hora), Distrito Federal - 2018 (parte 1)

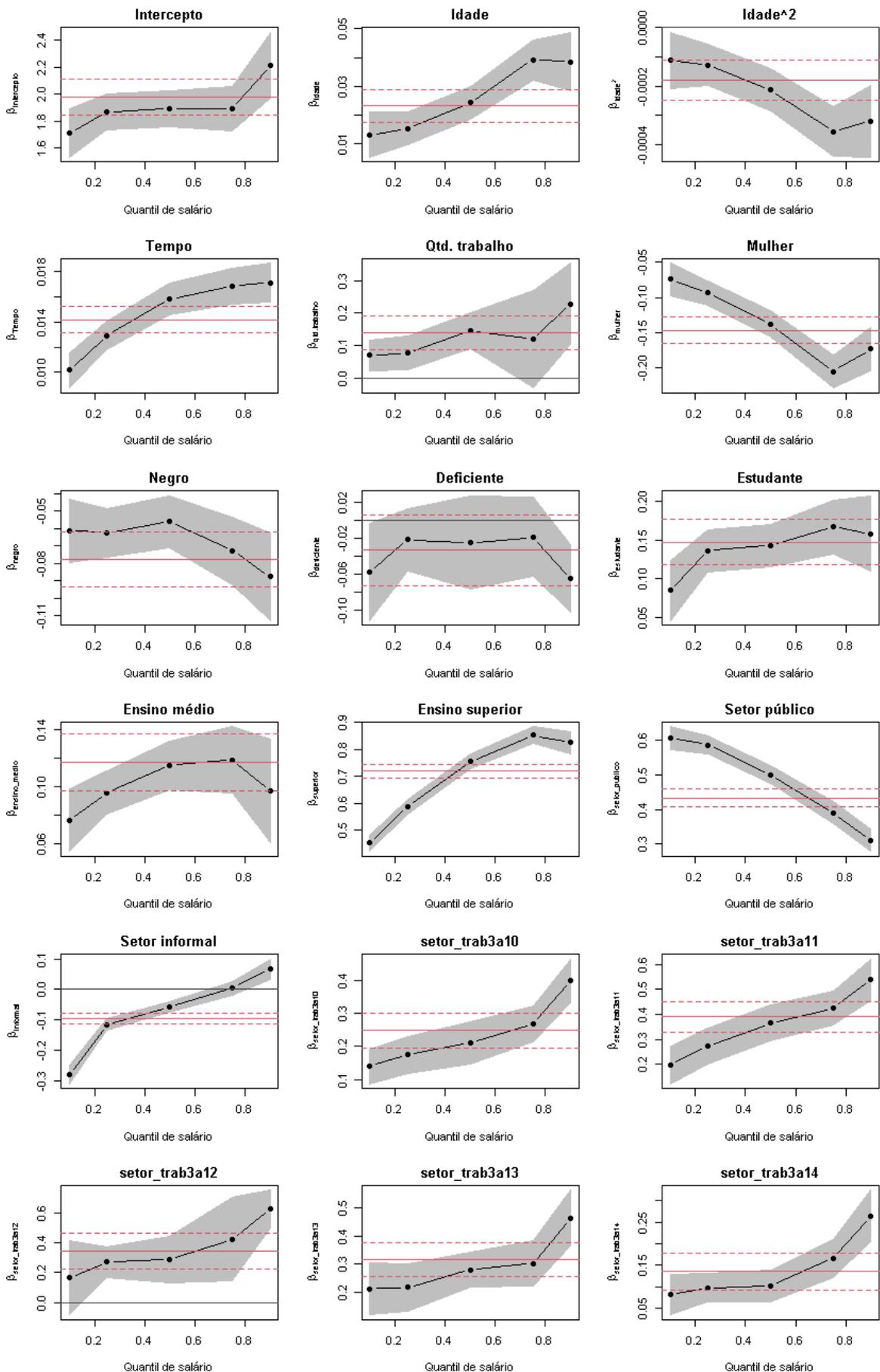


Gráfico 34: Estimativas e intervalos de confiança da regressão quantílica e da linear para Log(salário-hora), Distrito Federal - 2018 (parte final)

