

# Universidade Estadual do Oeste do Paraná

**Ciências Econômicas**

**Econometria II**

**Trabalho de heterocedasticidade – Questão 2**

Prof. Dra. Rosângela

Amanda Ricarte

Lucas Freire

Marcelo dos Santos

### **a) Apresente as estimativas do modelo.**

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2^2 + \delta_1 D_1 + \delta_2 D_2 + \delta_3 D_3 + \varepsilon_t$$

$y$  = Renda do trabalhador;

$\beta_0$  = Parâmetro que indica o intercepto do modelo;

$\beta_1$  = Parâmetro que indica o efeito marginal da idade sobre a renda do trabalhador;

$x_1$  = Variável independente discreta idade;

$\beta_2$  = Parâmetro que determina a curvatura da relação entre idade e renda do trabalhador;

$x_2^2$  = Variável independente polinomial discreta, da idade;

$\delta_1$  = Parâmetro que indica quanto o trabalhador residente da área urbana recebe a mais que o trabalhador residente da área rural;

$D_1$  = Variável dummie, 1 = trabalhador residente na área urbana, 0 = trabalhador residente da área rural;

$\delta_2$  = Parâmetro que indica quanto o trabalhador branco recebe a mais que o trabalhador preto;

$D_2$  = Variável dummie, 1 = trabalhador branco, 0 = trabalhador preto;

$\delta_3$  = Parâmetro que indica quanto o trabalhador homem recebe a mais que a trabalhadora mulher;

$D_3$  = Variável dummie, 1 = trabalhador homem, 0 = trabalhador mulher;

$\varepsilon_t$  = Termo de erro aleatório.

## a) Apresente as estimativas do modelo.

$$y = -2128,72 + 103,84x_1 - 1,05x_2^2 + 520,06D_1 + 595,49D_2 + 629,57D_3$$

OLS Regression Results						
=====						
Dep. Variable:	renda	R-squared:	0.035			
Model:	OLS	Adj. R-squared:	0.035			
Method:	Least Squares	F-statistic:	94.67			
Date:	Mon, 31 Mar 2025	Prob (F-statistic):	2.79e-98			
Time:	22:20:45	Log-Likelihood:	-1.2237e+05			
No. Observations:	13075	AIC:	2.448e+05			
Df Residuals:	13069	BIC:	2.448e+05			
Df Model:	5					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]
-----						
Intercept	-2128.7153	247.357	-8.606	0.000	-2613.571	-1643.859
idade	106.8395	11.419	9.356	0.000	84.457	129.222
I(idade ** 2)	-1.0527	0.141	-7.467	0.000	-1.329	-0.776
domic	520.0646	116.020	4.483	0.000	292.649	747.480
branco	595.4912	54.896	10.848	0.000	487.888	703.094
sexo	629.5673	49.729	12.660	0.000	532.091	727.044
=====						
Omnibus:	36407.049	Durbin-Watson:	1.808			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	2582868551.262			
Skew:	35.559	Prob(JB):	0.00			
Kurtosis:	2179.227	Cond. No.	1.87e+04			

**R<sup>2</sup> = 0,035**

O R<sup>2</sup> indica que apenas 4% da variação da variável dependente é explicada pelas variáveis independentes. Mesmo sendo baixo, o modelo ainda é válido (mas não satisfatório), se os coeficientes estimados forem estatisticamente significativos.

**Estatística F = 94,67**

O resultado para o teste F foi alto e satisfatório, além disso, seu p-valor foi estatisticamente significativo, nos permitindo rejeitar a hipótese nula de que todos os coeficientes são iguais a zero.

**Coefficientes**

Os resultados dos coeficientes estimados foram estatisticamente significativos. Entretanto, os valores para os erros padrão podem distorcer esses resultados. Nesse sentido, antes de analisar os parâmetros estimados, testaremos as hipóteses de heterocedasticidade presente no modelo, a fim de encontrar um modelo mais ajustado.

**b) Calcule o teste de Goldfeld-Quandt e analise seu resultado.**

$$H_0: \sigma_1^2 = \sigma_2^2$$
$$H_1: \sigma_1^2 \neq \sigma_2^2$$

```
Estatística F: 6.8373506436673335  
p-valor: 0.0
```

O teste GD é mais adequado para regressões simples. O resultado foi significativo, levando à rejeição da hipótese nula de que as variâncias estimadas na primeira parte são iguais às da segunda parte. Isso indica a presença de heterocedasticidade nos dados.

**c) Calcule o teste de Breusch-Pagan-Godfrey e analise seu resultado.**

$$H_0: \sigma_1^2 = \sigma_2^2$$
$$H_1: \sigma_1^2 \neq \sigma_2^2$$

```
Estatística Qui-quadrado: 1.7024975333098018  
p-valor: 0.8885849626866629
```

O teste BG é mais apropriado para regressões múltiplas, como esta, e base de dados grandes. O resultado do teste foi não significativo, então, não rejeitamos a hipótese nula de igualdade entre as variâncias estimadas, indicando homocedasticidade.

**Decisão dos pesquisadores:**

Considerando que o teste GQ deu significativo, vamos avançar para a transformação dos dados a fim de encontrar um melhor ajuste para o modelo.

**d) Caso seja necessário corrigir a heterocedasticidade do modelo, faça isso utilizando o “Modelo de Regressão com Erros-Padrão Robustos de White”.**

OLS Regression Results						
=====						
Dep. Variable:	renda	R-squared:	0.035			
Model:	OLS	Adj. R-squared:	0.035			
Method:	Least Squares	F-statistic:	110.0			
Date:	Mon, 31 Mar 2025	Prob (F-statistic):	3.73e-114			
Time:	22:10:54	Log-Likelihood:	-1.2237e+05			
No. Observations:	13075	AIC:	2.448e+05			
Df Residuals:	13069	BIC:	2.448e+05			
Df Model:	5					
Covariance Type:	HC0					
=====						
	coef	std err	z	P> z	[0.025	0.975]
-----						
Intercept	-2128.7153	261.014	-8.156	0.000	-2640.294	-1617.137
idade	106.8395	12.258	8.716	0.000	82.814	130.865
I(idade ** 2)	-1.0527	0.142	-7.392	0.000	-1.332	-0.774
domic	520.0646	66.868	7.777	0.000	389.005	651.124
branco	595.4912	45.340	13.134	0.000	506.626	684.356
sexo	629.5673	49.524	12.712	0.000	532.502	726.633
=====						
Omnibus:	36407.049	Durbin-Watson:	1.808			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	2582868551.262			
Skew:	35.559	Prob(JB):	0.00			
Kurtosis:	2179.227	Cond. No.	1.87e+04			

O método de White não altera os coeficientes estimados pelo modelo MQO. Ele recalcula os erros padrão dos coeficientes, levando em conta a heterocedasticidade presente nos resíduos, tornando os resultados do modelo mais confiáveis.

**e) Compare as duas estimativas em relação ao desvio-padrão dos parâmetros estimados.**

Variável	Coef. OLS	Std. OLS	p-valor OLS	Coef. white	Std. white	p-valor white
Intercept	-2128.715295	247.357094	8.434498e-18	-2128.715295	261.014224	3.475862e-16
idade	106.839474	11.418844	9.576527e-21	106.839474	12.257915	2.883092e-18
l(idade ** 2)	-1.052698	0.140982	8.729527e-14	-1.052698	0.142414	1.448638e-13
domic	520.064563	116.019608	7.438111e-06	520.064563	66.868245	7.399985e-15
branco	595.491215	54.895515	2.672382e-27	595.491215	45.340060	2.105648e-39
sexo	629.567298	49.729125	1.614843e-36	629.567298	49.524013	5.048461e-37

Após a aplicação do método de White, o desvio padrão do intercepto e variáveis independentes contínuas aumentaram, enquanto os desvios das variáveis binárias diminuíram. Isso significa que:

- O aumento nos erros padrão das variáveis contínuas reduziu a confiança na relação idade e renda;
- A redução nos erros padrão das binárias tornaram os coeficientes mais confiáveis, ou seja, os efeitos de urbanização, raça e gênero sobre a renda ficaram mais claros após a correção.

**f) Analise as estimativas do modelo, no que diz respeito: aos coeficientes, aos testes de hipóteses e ao Coeficiente de Determinação.**

OLS Regression Results						
=====						
Dep. Variable:	renda	R-squared:	0.035			
Model:	OLS	Adj. R-squared:	0.035			
Method:	Least Squares	F-statistic:	110.0			
Date:	Mon, 31 Mar 2025	Prob (F-statistic):	3.73e-114			
Time:	22:10:54	Log-Likelihood:	-1.2237e+05			
No. Observations:	13075	AIC:	2.448e+05			
Df Residuals:	13069	BIC:	2.448e+05			
Df Model:	5					
Covariance Type:	HC0					
=====						
	coef	std err	z	P> z	[0.025	0.975]
-----						
Intercept	-2128.7153	261.014	-8.156	0.000	-2640.294	-1617.137
idade	106.8395	12.258	8.716	0.000	82.814	130.865
I(idade ** 2)	-1.0527	0.142	-7.392	0.000	-1.332	-0.774
domic	520.0646	66.868	7.777	0.000	389.005	651.124
branco	595.4912	45.340	13.134	0.000	506.626	684.356
sexo	629.5673	49.524	12.712	0.000	532.502	726.633
=====						
Omnibus:	36407.049	Durbin-Watson:	1.808			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	2582868551.262			
Skew:	35.559	Prob(JB):	0.00			
Kurtosis:	2179.227	Cond. No.	1.87e+04			

**$R^2 = 0,035$**

O  $R^2$  não mudou após o ajuste dos desvios, mantendo a análise de que apenas 4% da variação da variável dependente é explicada pelas variáveis independentes. Mesmo sendo baixo, o modelo ainda é válido (mas não satisfatório), se os coeficientes estimados forem estatisticamente significativos.

**Estatística F = 110**

O resultado para o teste F foi maior que o primeiro e seu p-valor continua estatisticamente significativo, nos permitindo rejeitar a hipótese nula de que todos os coeficientes são iguais a zero.

**Coeficientes**

Todos os coeficientes apresentaram significância estatística, permitindo rejeitar a hipótese nula de que são iguais a zero.



f) Analise as estimativas do modelo, no que diz respeito: aos coeficientes, aos testes de hipóteses e ao Coeficiente de Determinação.

OLS Regression Results						
=====						
Dep. Variable:	renda	R-squared:	0.035			
Model:	OLS	Adj. R-squared:	0.035			
Method:	Least Squares	F-statistic:	110.0			
Date:	Mon, 31 Mar 2025	Prob (F-statistic):	3.73e-114			
Time:	22:10:54	Log-Likelihood:	-1.2237e+05			
No. Observations:	13075	AIC:	2.448e+05			
Df Residuals:	13069	BIC:	2.448e+05			
Df Model:	5					
Covariance Type:	HC0					
=====						
	coef	std err	z	P> z	[0.025	0.975]
-----						
Intercept	-2128.7153	261.014	-8.156	0.000	-2640.294	-1617.137
idade	106.8395	12.258	8.716	0.000	82.814	130.865
I(idade ** 2)	-1.0527	0.142	-7.392	0.000	-1.332	-0.774
domic	520.0646	66.868	7.777	0.000	389.005	651.124
branco	595.4912	45.340	13.134	0.000	506.626	684.356
sexo	629.5673	49.524	12.712	0.000	532.502	726.633
=====						
Omnibus:	36407.049	Durbin-Watson:	1.808			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	2582868551.262			
Skew:	35.559	Prob(JB):	0.00			
Kurtosis:	2179.227	Cond. No.	1.87e+04			

Coeficientes

**Intercepto:** representa o valor esperado da renda quando todas as variáveis explicativas são zero.

**Idade:** indica que a cada um ano a mais de idade, a renda do trabalhador aumenta em média R\$ 106,84 *ceteris paribus*.

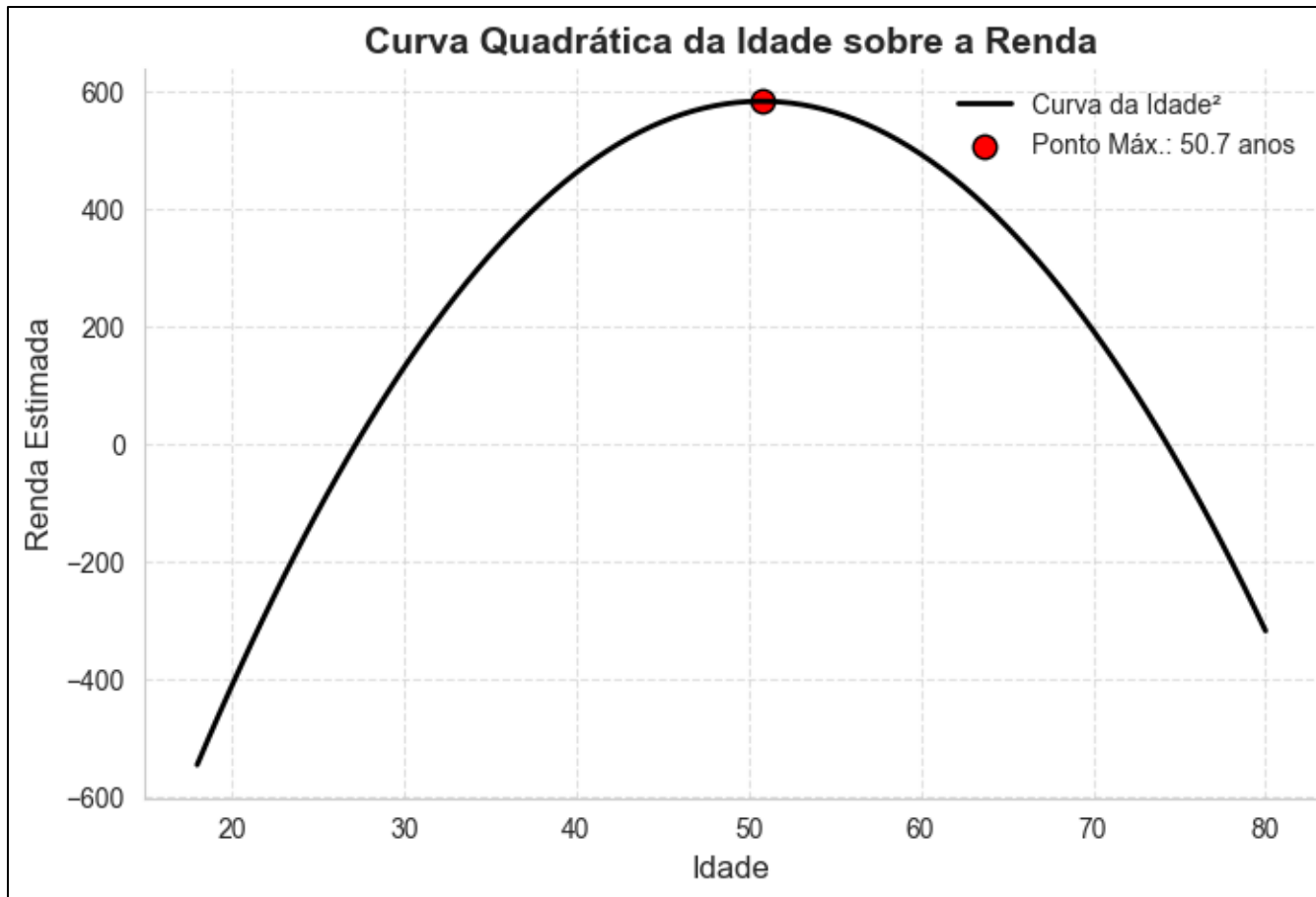
**Idade²:** embora a renda aumente com a idade, esse crescimento é cada vez menor, sugerindo um ponto de máximo na trajetória da renda ao longo da idade. Parábola concavidade para baixo.

**Domicílio urbano:** trabalhadores que moram em áreas urbanas ganham, em média, R\$ 520,06 a mais do que os que moram em áreas rurais.

**Raça branca:** trabalhadores brancos ganham, em média, R\$ 595,49 a mais do que os não brancos.

**Sexo masculino:** homens recebem, em média, R\$ 629,57 a mais do que mulheres, mantendo os demais fatores constantes.

**f) Analise as estimativas do modelo, no que diz respeito: aos coeficientes, aos testes de hipóteses e ao Coeficiente de Determinação.**



Embora a renda aumente com a idade, esse crescimento ocorre a uma taxa decrescente, indicando a existência de um ponto de máximo na trajetória da renda ao longo da vida profissional. A curva resultante tem concavidade para baixo, e esse ponto de máximo é estimado em aproximadamente 51 anos. Isso sugere que a renda do trabalhador tende a crescer até essa idade, a partir da qual começa a declinar, possivelmente refletindo a presença de etarismo no mercado de trabalho.

## **Repositório script**

<https://github.com/flucasbauer/heterocedasticidade-econometria>