

Universidade Estadual do Oeste do Paraná

Ciências Econômicas

Econometria II

Trabalho de heterocedasticidade – Questão 1

Prof. Dra. Rosângela

Amanda Ricarte

Lucas Freire

Marcelo dos Santos

a) Apresente as estimativas do modelo.

$$y = \beta_0 + \beta_1 x_1 + \varepsilon_t$$

y = Renda do trabalhador;

β_0 = Parâmetro que indica o intercepto do modelo;

β_1 = Parâmetro que indica o efeito marginal da idade sobre a renda do trabalhador;

x_1 = Variável independente discreta idade.

ε_t = Termo de erro aleatório.

a) Apresente as estimativas do modelo.

$y = 613,65 + 23,97x_1$

OLS Regression Results						
=====						
Dep. Variable:	renda	R-squared:	0.011			
Model:	OLS	Adj. R-squared:	0.011			
Method:	Least Squares	F-statistic:	142.1			
Date:	Mon, 31 Mar 2025	Prob (F-statistic):	1.36e-32			
Time:	15:57:04	Log-Likelihood:	-1.2253e+05			
No. Observations:	13075	AIC:	2.451e+05			
Df Residuals:	13073	BIC:	2.451e+05			
Df Model:	1					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

Intercept	613.6543	78.073	7.860	0.000	460.620	766.689
idade	23.9690	2.011	11.921	0.000	20.028	27.910
=====						
Omnibus:	35903.355	Durbin-Watson:	1.805			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	2314184423.294			
Skew:	34.270	Prob(JB):	0.00			
Kurtosis:	2062.887	Cond. No.	122.			
=====						

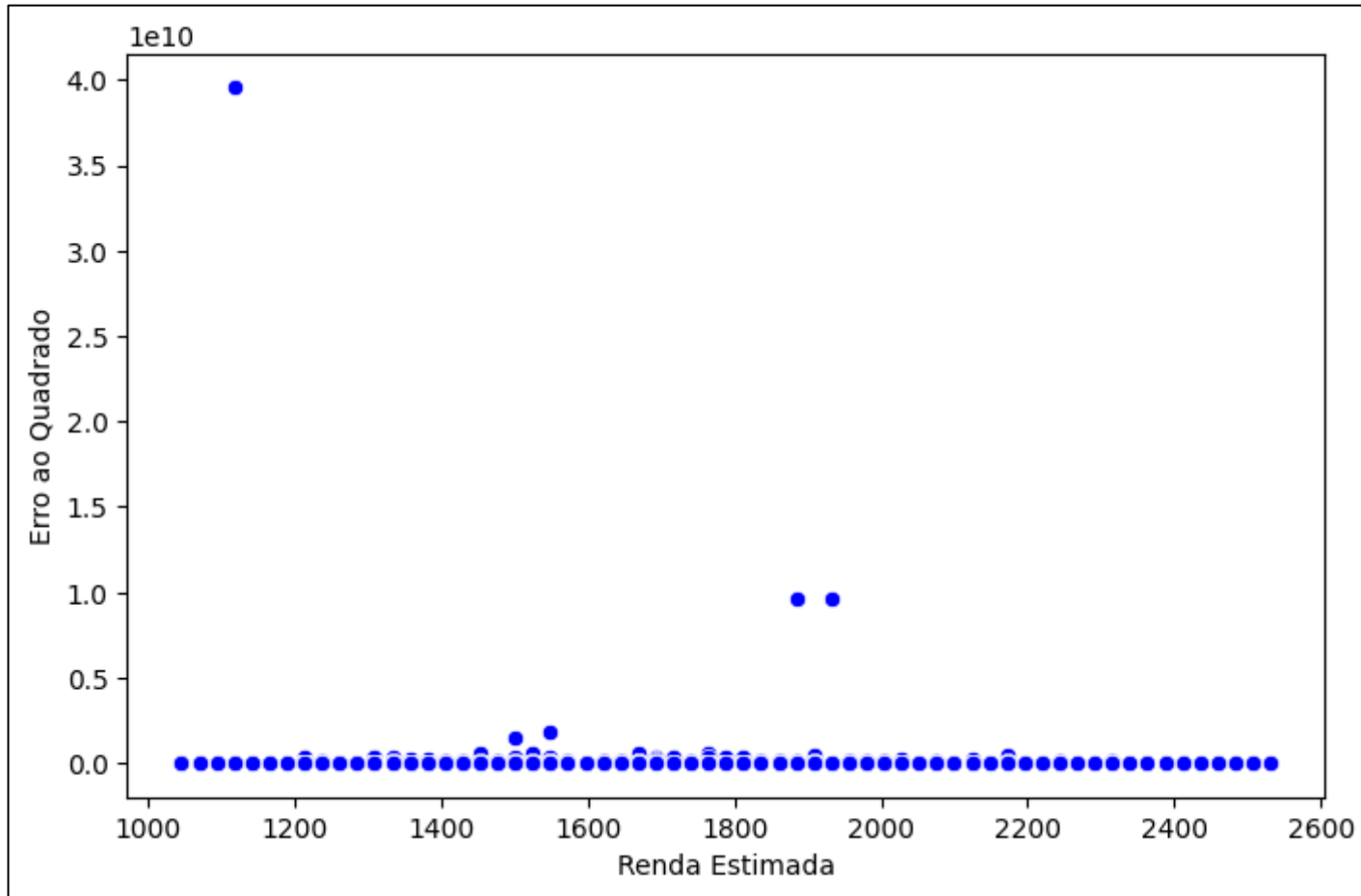
R² = 0,011

O R² indica que apenas 1% da variação da variável dependente é explicada pela variável independente. Mesmo sendo baixo, o modelo ainda é válido (mas não satisfatório), se os coeficientes estimados forem estatisticamente significativos.

Coeficientes

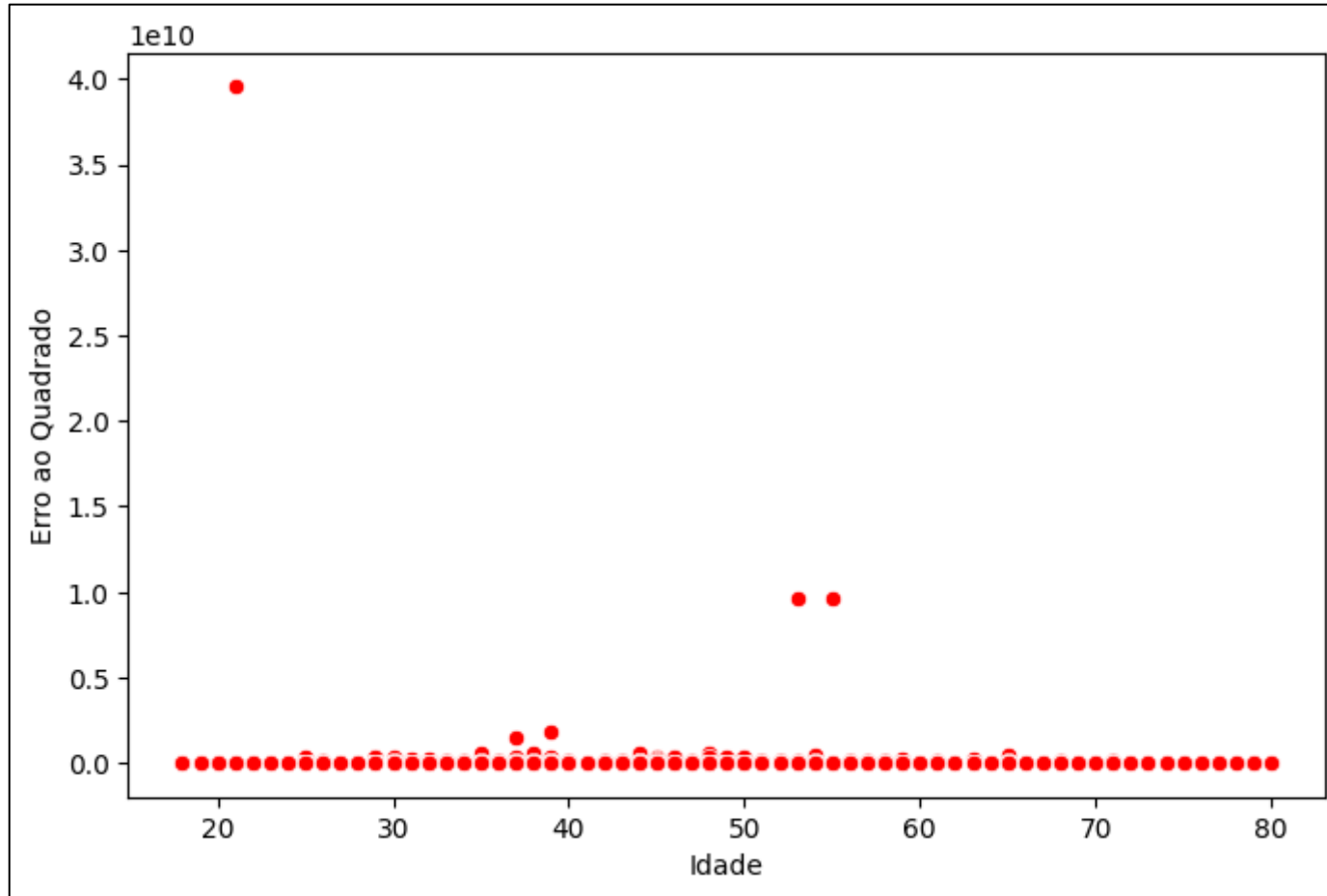
Antes de analisar os coeficientes, vamos aplicar os testes da hipótese da presença de heterocedasticidade.

b) Plote o gráfico do Erro ao quadrado em função da Renda Estimada e explique se, de acordo com o observado, é possível afirmar que existe heterocedasticidade.



Os pontos no gráfico ficaram alinhados indicando que não há um padrão claro de heterocedasticidade da variável explicativa renda. Os dados podem possuir resíduos de variância constante. No entanto, prosseguiremos com os teste para comprovar esta hipótese.

c) Plote o gráfico do Erro ao quadrado em função da Idade e explique se, de acordo com o observado, é possível afirmar que existe heterocedasticidade.



Os pontos no gráfico ficaram alinhados indicando que não há um padrão claro de heterocedasticidade da variável explicativa idade. Os dados podem possuir resíduos de variância constante. No entanto, prosseguiremos com os teste para comprovar esta hipótese.

d) 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: 6.640786130942513  
p-valor: 0.0
```

O teste GD é especialmente utilizado em regressões simples, como neste caso. 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, diferentemente da interpretação dos gráficos anteriores.

e) 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: 0.0014927161965938573
p-valor: 0.9691808505818943

O teste BG é mais apropriado para regressões múltiplas e amostras grandes. Como o resultado não foi significativo, não rejeitamos a hipótese nula de igualdade entre as variâncias estimadas, indicando homoscedasticidade.

Decisão dos pesquisadores:

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

f) Caso seja necessário corrigir a heterocedasticidade do modelo, faça isso utilizando a variável Idade como a base para a correção.

```
=====
                        WLS Regression Results
=====
Dep. Variable:          renda      R-squared:                0.013
Model:                  WLS        Adj. R-squared:             0.013
Method:                 Least Squares    F-statistic:           166.9
Date:                   Mon, 31 Mar 2025    Prob (F-statistic):    6.01e-38
Time:                   16:48:04      Log-Likelihood:       -1.2324e+05
No. Observations:      13075        AIC:                  2.465e+05
Df Residuals:          13073        BIC:                  2.465e+05
Df Model:               1
Covariance Type:       nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
const	437.6729	77.335	5.659	0.000	286.086	589.260
idade	28.7500	2.225	12.919	0.000	24.388	33.112

```
=====
Omnibus:                41677.484    Durbin-Watson:           1.847
Prob(Omnibus):           0.000      Jarque-Bera (JB):       9290669134.425
Skew:                    51.453      Prob(JB):               0.00
Kurtosis:                4131.320    Cond. No.               105.
=====
```

	Coef. OLS	Desvio-Padrão OLS	Coef. WLS	Desvio-Padrão WLS
const	613.654283	78.073018	437.672910	77.334619
idade	23.968972	2.010646	28.750011	2.225433

Após o ajuste do modelo identificamos que:

- os coeficientes continuam estatisticamente significativos;
- O desvio padrão da constante diminuiu (ótimo sinal); e o
- Desvio padrão da idade aumentou (sinal ruim).

O modelo ponderado ajustou melhor a heterocedasticidade dos erros, reduzindo o desvio padrão da constante, o que indica menor incerteza nessa estimativa. No entanto, isso ocorreu à custa de um aumento no desvio padrão do coeficiente da idade, aumentando a variabilidade dessa estimativa.

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

	Coef. OLS	Desvio-Padrão OLS	Coef. WLS	Desvio-Padrão WLS
const	613.654283	78.073018	437.672910	77.334619
idade	23.968972	2.010646	28.750011	2.225433

Como os coeficientes foram estatisticamente significativos tanto no modelo antes do ajuste quanto após o ajuste, **optamos por analisar os parâmetros estimados pelo primeiro modelo.** Nossa decisão foi baseada no desvio padrão do coeficiente de idade, pois consideramos a confiabilidade desse resultado o mais relevante para a análise.

g) 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:	0.011			
Model:	OLS	Adj. R-squared:	0.011			
Method:	Least Squares	F-statistic:	142.1			
Date:	Mon, 31 Mar 2025	Prob (F-statistic):	1.36e-32			
Time:	15:57:04	Log-Likelihood:	-1.2253e+05			
No. Observations:	13075	AIC:	2.451e+05			
Df Residuals:	13073	BIC:	2.451e+05			
Df Model:	1					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

Intercept	613.6543	78.073	7.860	0.000	460.620	766.689
idade	23.9690	2.011	11.921	0.000	20.028	27.910
=====						
Omnibus:	35903.355	Durbin-Watson:	1.805			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	2314184423.294			
Skew:	34.270	Prob(JB):	0.00			
Kurtosis:	2062.887	Cond. No.	122.			
=====						

R² = 0,011

O R² indica que apenas 1% da variação da variável dependente é explicada pela variável independente. Mesmo sendo baixo, o modelo ainda é válido (mas não satisfatório), se os coeficientes estimados forem estatisticamente significativos.

Coeficientes

Intercepto: Estatisticamente significativo. Quando a idade é zero, a renda prevista seria R\$ 613,65. No entanto, esse valor pode não ter interpretação prática, pois a idade zero não faz sentido no contexto econômico.

Idade: Estatisticamente significativo. Para cada aumento de 1 ano na idade, a renda média aumenta em R\$ 23,97, mantendo todas as outras condições constantes.

Repositório script

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