

ACA228 - Modelos de Regressão e Previsão

Regressão Linear Multipla: Verificando as hipóteses I

Prof. Carlos Trucíos
carlos.trucios@facc.ufrj.br
ctruciosm.github.io

Faculdade de Administração e Ciências Contábeis,
Universidade Federal do Rio de Janeiro

Aula 13

Introdução

Tipos de resíduos

Gráfico de resíduos

Heterocedasticidade

Introdução

Introdução

O MRLM é contruido sob as seguintes hipóteses:

- ▶ HRLM1: Modelo Populacional Linear nos parâmetros
- ▶ HRLM2: Amostragem aleatória
- ▶ HRLM3: Colinearidade não perfeita
- ▶ HRLM4: Média condicional zero
- ▶ HRLM5: Variância constante
- ▶ HRLM6: Normalidade

O que acontece se algumas dessas hipóteses não são verificadas?

Introdução

A não verificação das hipóteses pode levar a resultados instáveis, *i.e.* uma amostra diferente pode levar a resultados totalmente diferentes e inclusive conclusões opostas.

Introdução

A não verificação das hipóteses pode levar a resultados instáveis, *i.e.* uma amostra diferente pode levar a resultados totalmente diferentes e inclusive conclusões opostas.

- ▶ Os resultados padrão do ajuste do modelo como R^2 , Teste-t, Teste-F não nos permitem saber se alguma das hipóteses foi violada.

Introdução

A não verificação das hipóteses pode levar a resultados instáveis, *i.e.* uma amostra diferente pode levar a resultados totalmente diferentes e inclusive conclusões opostas.

- ▶ Os resultados padrão do ajuste do modelo como R^2 , Teste-t, Teste-F não nos permitem saber se alguma das hipóteses foi violada.
- ▶ Os resíduos nos ajudarão nesta tarefa

$$\hat{u}_i = y_i - \hat{y}_i$$

Introdução

A não verificação das hipóteses pode levar a resultados instáveis, *i.e.* uma amostra diferente pode levar a resultados totalmente diferentes e inclusive conclusões opostas.

- ▶ Os resultados padrão do ajuste do modelo como R^2 , Teste-t, Teste-F não nos permitem saber se alguma das hipóteses foi violada.
- ▶ Os resíduos nos ajudarão nesta tarefa

$$\hat{u}_i = y_i - \hat{y}_i$$

- ▶ A análise dos resíduos é uma maneira eficaz de descobrir vários tipos de *inadequações* no modelo ajustado.

Tipos de resíduos

Tipos de resíduos

Standardized Residuals:

Sabemos que

$$\hat{u} = y - \underbrace{\hat{y}}_{X'\hat{\beta}} = y - X' \underbrace{(X'X)^{-1}X'y}_{\hat{\beta}} = (I - \underbrace{X'(X'X)^{-1}X'}_H)y = (I - H)y$$

vamos *padronizar* cada resíduo pelo seu respectivo desvio padrão

Tipos de resíduos

Standardized Residuals:

Sabemos que

$$\hat{u} = y - \underbrace{\hat{y}}_{X'\hat{\beta}} = y - X' \underbrace{(X'X)^{-1}X'y}_{\hat{\beta}} = (I - \underbrace{X'(X'X)^{-1}X'}_H)y = (I - H)y$$

vamos *padronizar* cada resíduo pelo seu respectivo desvio padrão

$$V(\hat{u}_i|X) = V(\underbrace{[0 \dots 1 \dots 0]}_{\mathbb{I}_i} \hat{u}|X) = \mathbb{I}_i \underbrace{V(\hat{u}|X)}_{(I-H)\sigma^2} \mathbb{I}_i' = \sigma^2(1 - h_{ii})$$

Tipos de resíduos

Standardized Residuals:

Sabemos que

$$\hat{u} = y - \underbrace{\hat{y}}_{X'\hat{\beta}} = y - X' \underbrace{(X'X)^{-1}X'y}_{\hat{\beta}} = (I - \underbrace{X'(X'X)^{-1}X'}_H)y = (I - H)y$$

vamos *padronizar* cada resíduo pelo seu respectivo desvio padrão

$$V(\hat{u}_i|X) = V(\underbrace{[0 \dots 1 \dots 0]}_{\mathbb{I}_i} \hat{u}|X) = \mathbb{I}_i \underbrace{V(\hat{u}|X)}_{(I-H)\sigma^2} \mathbb{I}_i' = \sigma^2(1 - h_{ii})$$

$$r_i = \frac{\hat{u}_i}{\sqrt{\hat{\sigma}^2(1 - h_{ii})}}$$

Tipos de resíduos

Studentized Residuals

Seja $\hat{y}_{(i)}$ o valor estimado de y_i utilizando todas as observações menos a i -ésima.

$$\hat{u}_{(i)} = y_i - \hat{y}_{(i)} = \frac{\hat{u}_i}{1 - h_{ii}}$$

Se dividirmos pelo desvio padrão de $\hat{u}_{(i)}$, temos:

$$t_i = \frac{\hat{u}_i}{\sqrt{S_{(i)}^2(1 - h_{ii})}} = r_i \sqrt{\frac{n - k - 2}{n - k - 1 - r_i^2}}$$

onde $S_{(i)}^2 = \frac{(n-k-1)\hat{\sigma}^2 - \hat{u}_i^2/(1-h_{ii})}{n-k-2}$

Tipos de resíduos

Resíduo	Notação	Formula
Standardized	r_i	$\frac{\hat{u}_i}{\sqrt{\hat{\sigma}^2(1 - h_{ii})}}$
Studentized	t_i	$\frac{\hat{u}_i}{\sqrt{S_{(i)}^2(1 - h_{ii})}}$

Studentized é preferido.

Tipos de resíduos

Resíduo	Notação	Formula
Standardized	r_i	$\frac{\hat{u}_i}{\sqrt{\hat{\sigma}^2(1 - h_{ii})}}$
Studentized	t_i	$\frac{\hat{u}_i}{\sqrt{S_{(i)}^2(1 - h_{ii})}}$

Studentized é preferido.

Com os resíduos vamos identificar **outliers** ou **valores extremos** e as hipóteses do **modelo linear clássico**

Tipos de resíduos

```
library(wooldridge)
modelo = lm(log(bwght) ~ npvis + I(npvis^2), data = bwght2)
uhat = residuals(modelo)
r = rstandard(modelo)
t = rstudent(modelo)
```


Gráfico de resíduos

Gráfico para detecção de outliers/extremos

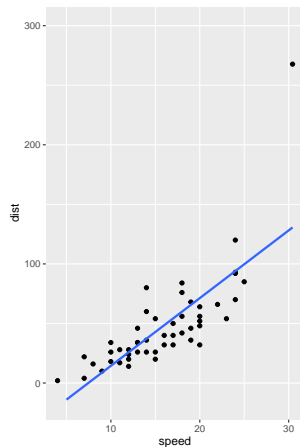
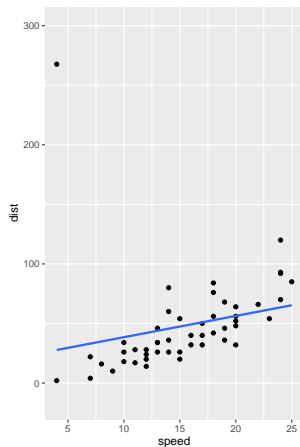
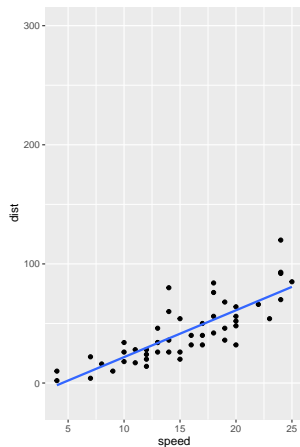


Gráfico para detecção de outliers/extremos: residuais vs \hat{y}

- ▶ Faremos o gráfico dos valores estimados \hat{y} (eixo X) vs. os residuais (eixo Y)
- ▶ Se algum ponto está muito afastado dos outros, ele será um possível outliers
- ▶ Faremos os gráficos utilizando os três tipos de resíduos

Mas os *Studentized Residuals* são preferidos quando fazemos análise de resíduos.

Gráfico para detecção de outliers/extremos: residuais vs yhat

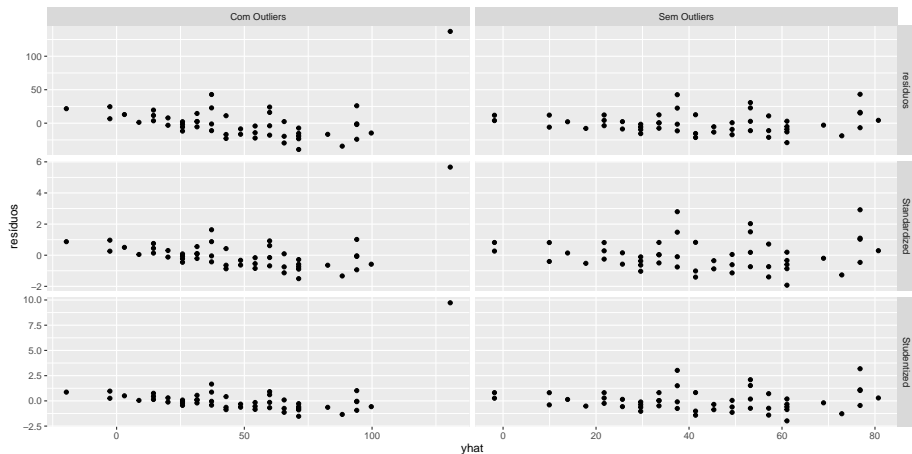


Gráfico de probabilidade Normal

- ▶ Quando precisamos verificar a hipótese de normalidade
- ▶ Nos permite identificar se as caudas são mais pesadas do que as caudas da distribuição Normal
- ▶ Nos permite identificar assimetrias
- ▶ Sob Normalidade, esperamos que os pontos estejam sobre a reta

```
qqnorm(normal)  
qqline(normal)
```

Gráfico de probabilidade Normal

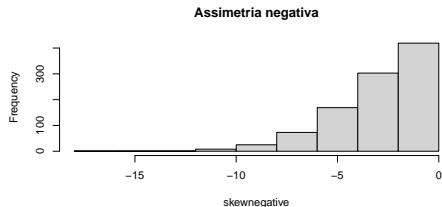
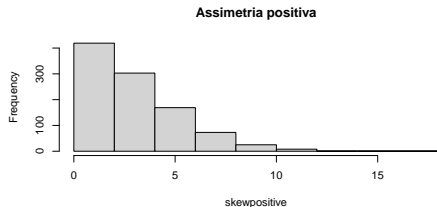
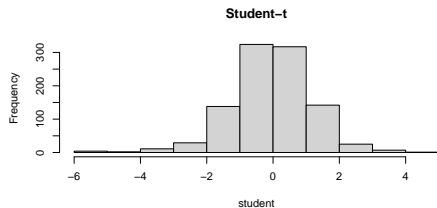
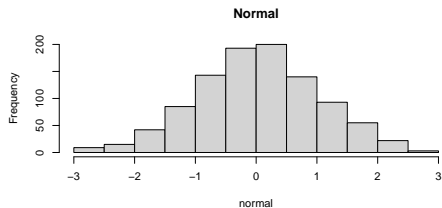


Gráfico de probabilidade Normal

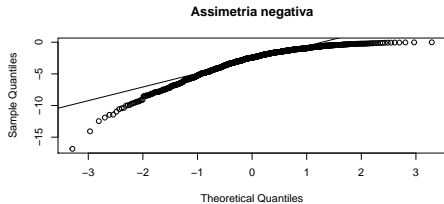
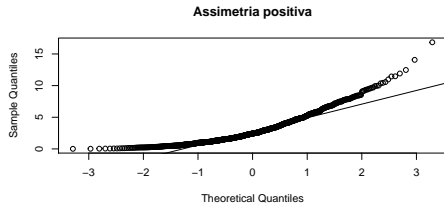
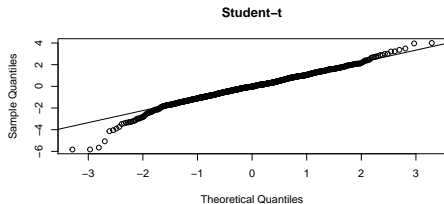
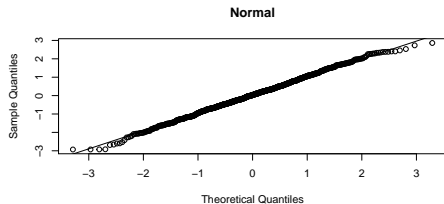
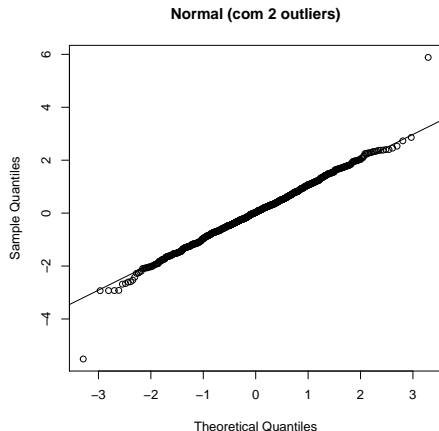
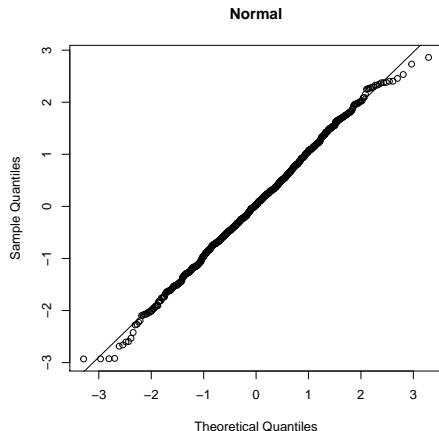
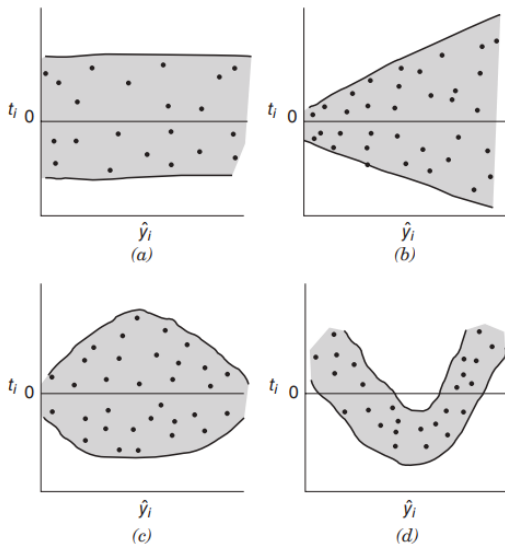


Gráfico de probabilidade Normal: Detecção de outliers



Gráficos para verificar homecasticidade e forma funcional



Gráficos para verificar homecasticidade e forma funcional

Preferivelmente utilizamos os *Studentized Residuals*

$$t_i = \frac{\hat{u}_i}{\sqrt{S_{(i)}^2(1 - h_{ii})}},$$

onde $S_{(i)}^2 = \frac{(n-p)\hat{\sigma}^2 - \hat{u}_i^2/(1-h_{ii})}{n-p-1}$

- ▶ (a): modelo aparentemente ok
- ▶ (b) e (c): indicam que a variância dos erros não é constante
- ▶ (d): não linearidade

Cuidado!

O grafico é dos residuos vs \hat{y} (não y)

Gráfico de sequencia dos resíduos e ACF

- ▶ Até agora vimos gráficos que nos ajudam a detectar heterocedasticidade, não linearidade, normalidade.

Gráfico de sequencia dos resíduos e ACF

- ▶ Até agora vimos gráficos que nos ajudam a detectar heterocedasticidade, não linearidade, normalidade.
- ▶ Agora veremos dois gráficos que nos darão uma ideia sobre a correlação dos resíduos

Gráfico de sequencia dos resíduos e ACF

- ▶ Até agora vimos gráficos que nos ajudam a detectar heterocedasticidade, não linearidade, normalidade.
- ▶ Agora veremos dois gráficos que nos darão uma ideia sobre a correlação dos resíduos
- ▶ Gráfico de sequência (série temporal) e função de autocorrelação

Gráfico de sequencia dos resíduos e ACF

- ▶ Até agora vimos gráficos que nos ajudam a detectar heterocedasticidade, não linearidade, normalidade.
- ▶ Agora veremos dois gráficos que nos darão uma ideia sobre a correlação dos resíduos
- ▶ Gráfico de sequência (série temporal) e função de autocorrelação

Gráfico de sequencia dos resíduos e ACF

- ▶ Até agora vimos gráficos que nos ajudam a detectar heterocedasticidade, não linearidade, normalidade.
- ▶ Agora veremos dois gráficos que nos darão uma ideia sobre a correlação dos resíduos
- ▶ Gráfico de sequência (série temporal) e função de autocorrelação

Lembre-se, sob HMRLM2, os erros são não correlacionados (na verdade eles são independentes), então esperamos também que os resíduos sejam não correlacionados.

Gráfico de sequencia dos resíduos e ACF

```
ts.plot(normal)
```

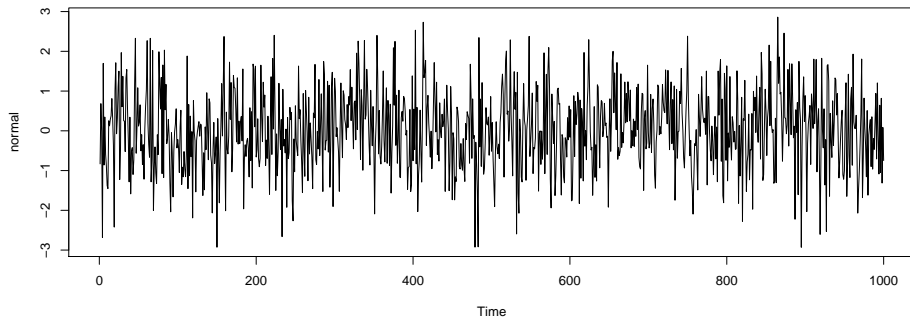
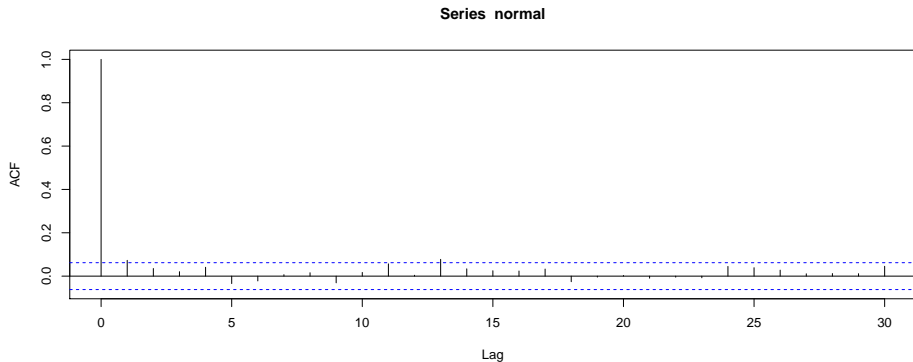


Gráfico de sequencia dos resíduos e ACF

```
acf(normal)
```



O que fazer após ajustar o modelo?

1. Gráfico de sequência e ACF dos resíduos
2. Gráfico dos resíduos vs os valores ajustados \hat{y}
3. Gráfico de probabilidade normal*
4. Verificar se existem outliers ou valores extremos

Heterocedasticidade

Homocedasticidade

- ▶ Significa que a variância do erro, condicional nas variáveis explicativas, é constante *i.e.*

$$V(u|X) = \sigma^2 I$$

- ▶ É necessária para definir o *teste t*, *teste F*, *IC para β*
- ▶ Sua ausência tem consequências no método MQO
- ▶ Heterocedasticidade: não homocedasticidade

Heterocedasticidade

Homocedasticidade

- ▶ Significa que a variância do erro, condicional nas variáveis explicativas, é constante *i.e.*

$$V(u|X) = \sigma^2 I$$

- ▶ É necessária para definir o *teste t*, *teste F*, *IC para β*
- ▶ Sua ausência tem consequências no método MQO
- ▶ Heterocedasticidade: não homocedasticidade

Discutiremos as consequências, detecção e soluções ao problema de Heterocedasticidade.

Heterocedasticidade

A falta de homocedasticidade afeta $V(\hat{\beta})$

- ▶ Os intervalos de confiança (para β) construídos não são mais válidos

Heterocedasticidade

A falta de homocedasticidade afeta $V(\hat{\beta})$

- ▶ Os intervalos de confiança (para β) construídos não são mais válidos
- ▶ Os testes de hipóteses não são mais válidos

Heterocedasticidade

A falta de homocedasticidade afeta $V(\hat{\beta})$

- ▶ Os intervalos de confiança (para β) construídos não são mais válidos
- ▶ Os testes de hipóteses não são mais válidos

Heterocedasticidade

A falta de homocedasticidade afeta $V(\hat{\beta})$

- ▶ Os intervalos de confiança (para β) construídos não são mais válidos
- ▶ Os testes de hipóteses não são mais válidos

As estatísticas utilizadas para testar hipóteses não são mais válidas na presença de heterocedasticidade

Heterocedasticidade

A falta de homocedasticidade afeta $V(\hat{\beta})$

- ▶ Os intervalos de confiança (para β) construídos não são mais válidos
- ▶ Os testes de hipóteses não são mais válidos

As estatísticas utilizadas para testar hipóteses não são mais válidas na presença de heterocedasticidade

Na presença de heterocedasticidade, precisamos corrigir a estimação de $V(\hat{\beta})$, para isto utilizaremos o estimador de White

Heterocedasticidade

Estimador de White

Seja $y_i = \beta_0 + \beta_1 x_i + u_i$, então o estimator MQO é da forma

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x}) u_i}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Sob heterocedasticidade, *i.e.* $V(u_i|X) = \sigma_i^2$, temos que:

$$V(\hat{\beta}_1|X) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \sigma_i^2}{\left[\sum_{i=1}^n (x_i - \bar{x})^2 \right]^2}$$

Heterocedasticidade

Estimador de White

Como σ_i^2 é desconhecido, estimamos $V(\hat{\beta}_1|X) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \sigma_i^2}{\left[\sum_{i=1}^n (x_i - \bar{x})^2\right]^2}$ por

Estimador de White | White-Huber

$$\hat{V}(\hat{\beta}_1|X) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \hat{u}_i^2}{\left[\sum_{i=1}^n (x_i - \bar{x})^2\right]^2}$$

Heterocedasticidade

No modelo $y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u$

Sob HRLM1–HRLM4 e $V(u|X) = E(uu'|X) = \Omega$,

Estimador de White | White-Huber

$$\widehat{V}(\hat{\beta}|X) = (X'X)^{-1}X'\widehat{\Omega}X(X'X)^{-1}$$

onde $\widehat{\Omega} = \text{diag}\{\hat{u}_1^2, \dots, \hat{u}_n^2\}$

Heterocedasticidade

No modelo $y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u$

Sob HRLM1–HRLM4 e $V(u|X) = E(uu'|X) = \Omega$,

Estimador de White | White-Huber

$$\widehat{V}(\hat{\beta}|X) = (X'X)^{-1}X'\widehat{\Omega}X(X'X)^{-1}$$

onde $\widehat{\Omega} = \text{diag}\{\hat{u}_1^2, \dots, \hat{u}_n^2\}$

Usando as variâncias obtidas pelo estimador de White, os *testes t*, *testes F* e *IC* podem ser obtidos como de costume.

Heterocedasticidade: wage1

```
modelo = lm(log(wage) ~ homemcasado + mulhercasada +  
             mulhersolteira + educ + exper +  
             I(exper^2) + tenure + I(tenure^2),  
             data = wage1)  
round(coef(modelo),4)
```

##	(Intercept)	homemcasado	mulhercasada	mulhersolteira
##	0.3214	0.2127	-0.1983	-0.1104
##	exper	I(exper^2)	tenure	I(tenure^2)
##	0.0268	-0.0005	0.0291	-0.0005

Heterocedasticidade: wage1

```
library(sandwich); library(lmtest)
coeftest(modelo,vcov= vcovHC(modelo, type="HC"))
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.32137810  0.10852842   2.9612 0.0032049 **
## homemcasado   0.21267568  0.05665095   3.7541 0.0001937 ***
## mulhercasada -0.19826760  0.05826505  -3.4029 0.0007186 ***
## mulhersolteira -0.11035021  0.05662551  -1.9488 0.0518632 .
## educ          0.07891028  0.00735096  10.7347 < 2.2e-16 ***
## exper         0.02680057  0.00509497   5.2602 2.111e-07 ***
## I(exper^2)    -0.00053525  0.00010543  -5.0770 5.360e-07 ***
## tenure       0.02908752  0.00688128   4.2270 2.800e-05 ***
## I(tenure^2)   -0.00053314  0.00024159  -2.2068 0.0277670 *
```

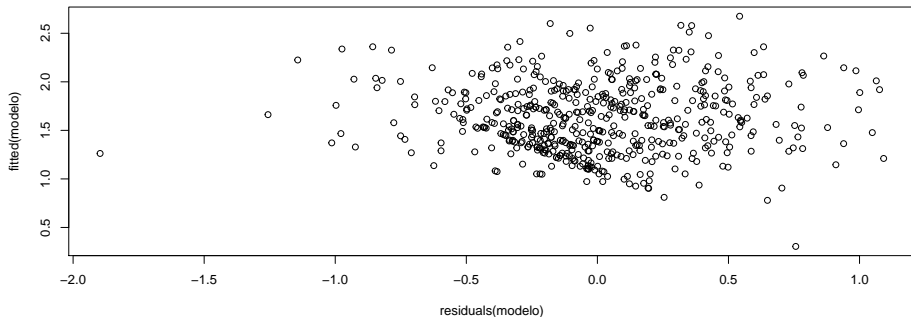

Heterocedasticidade: wage1

Comparação

##	betas	Std	White	P-valor	P-valor_W
## (Intercept)	0.32138	0.10001	0.10853	0.00139	0.00320
## homemcasado	0.21268	0.05536	0.05665	0.00014	0.00019
## mulhercasada	-0.19827	0.05784	0.05827	0.00066	0.00072
## mulhersolteira	-0.11035	0.05574	0.05663	0.04827	0.05186
## educ	0.07891	0.00669	0.00735	0.00000	0.00000
## exper	0.02680	0.00524	0.00509	0.00000	0.00000
## I(exper^2)	-0.00054	0.00011	0.00011	0.00000	0.00000
## tenure	0.02909	0.00676	0.00688	0.00002	0.00003
## I(tenure^2)	-0.00053	0.00023	0.00024	0.02153	0.02777

Heterocedasticidade: wage1

```
plot(residuals(modelo),fitted(modelo))
```



Heterocedasticidade: gpa3

```
library(dplyr)
newgpa = gpa3 %>% filter(term == 2)
modelo = lm(cumgpa~sat+hsperc + tothrs +
            female + black+white, data = newgpa)
round(summary(modelo)$coef,6)
```

##		Estimate	Std. Error	t value	Pr(> t)
##	(Intercept)	1.470065	0.229803	6.397063	0.000000
##	sat	0.001141	0.000179	6.388504	0.000000
##	hsperc	-0.008566	0.001240	-6.906004	0.000000
##	tothrs	0.002504	0.000731	3.425511	0.000685
##	female	0.303433	0.059020	5.141166	0.000000
##	black	-0.128284	0.147370	-0.870486	0.384616
##	white	-0.058722	0.140990	-0.416497	0.677295

Heterocedasticidade: gpa3

```
coeftest(modelo,vcov= vcovHC(modelo, type="HC"))
```

```
##
```

```
## t test of coefficients:
```

```
##
```

##		Estimate	Std. Error	t value	Pr(> t)	
##	(Intercept)	1.47006477	0.21855969	6.7261	6.888e-11	***
##	sat	0.00114073	0.00018969	6.0136	4.468e-09	***
##	hsperc	-0.00856636	0.00140430	-6.1001	2.744e-09	***
##	tothrs	0.00250400	0.00073353	3.4136	0.0007142	***
##	female	0.30343329	0.05856959	5.1807	3.693e-07	***
##	black	-0.12828368	0.11809549	-1.0863	0.2780880	
##	white	-0.05872173	0.11032164	-0.5323	0.5948631	

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Heterocedasticidade: gpa3

Comparação

##	betas	Std	White	P-valor	P-valor_W
## (Intercept)	1.47006	0.22980	0.21856	0.00000	0.00000
## sat	0.00114	0.00018	0.00019	0.00000	0.00000
## hsperc	-0.00857	0.00124	0.00140	0.00000	0.00000
## tothrs	0.00250	0.00073	0.00073	0.00068	0.00071
## female	0.30343	0.05902	0.05857	0.00000	0.00000
## black	-0.12828	0.14737	0.11810	0.38462	0.27809
## white	-0.05872	0.14099	0.11032	0.67730	0.59486

Heterocedasticidade: gpa3

Seja

$$H_0 : \beta_{black} = 0, \quad \beta_{white} = 0 \quad \text{vs.} \quad H_1 : H_0 \quad \text{n\~{a}o \acute{e} verdade}$$

- Precisamos do Teste F

Heterocedasticidade: gpa3

Seja

$$H_0 : \beta_{black} = 0, \quad \beta_{white} = 0 \quad \text{vs.} \quad H_1 : H_0 \quad \text{não é verdade}$$

- ▶ Precisamos do Teste F
- ▶ Modelo irrestrito

$$cumgpa = \beta_0 + \beta_1 sat + \beta_2 hsperc + \beta_3 tothrs + \beta_4 female + \beta_5 black + \beta_6 white + u$$

Heterocedasticidade: gpa3

Seja

$$H_0 : \beta_{black} = 0, \quad \beta_{white} = 0 \quad \text{vs.} \quad H_1 : H_0 \quad \text{não é verdade}$$

- ▶ Precisamos do Teste F
- ▶ Modelo irrestrito

$$cumgpa = \beta_0 + \beta_1 sat + \beta_2 hsperc + \beta_3 tothrs + \beta_4 female + \beta_5 black + \beta_6 white + u$$

- ▶ Modelo restrito

$$cumgpa = \beta_0 + \beta_1 sat + \beta_2 hsperc + \beta_3 tothrs + \beta_4 female + u$$

Heterocedasticidade: gpa3

```
modeloi = lm(cumgpa~sat+hsperc + tothrs + female + black +  
             white, data = newgpa)  
modelor = lm(cumgpa~sat+hsperc + tothrs + female, data = newgpa)  
# Teste F clássico  
anova(modelor,modeloi)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: cumgpa ~ sat + hsperc + tothrs + female
```

```
## Model 2: cumgpa ~ sat + hsperc + tothrs + female + black + wh
```

```
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
```

```
## 1      361 79.362
```

```
## 2      359 79.062  2    0.29934 0.6796 0.5075
```

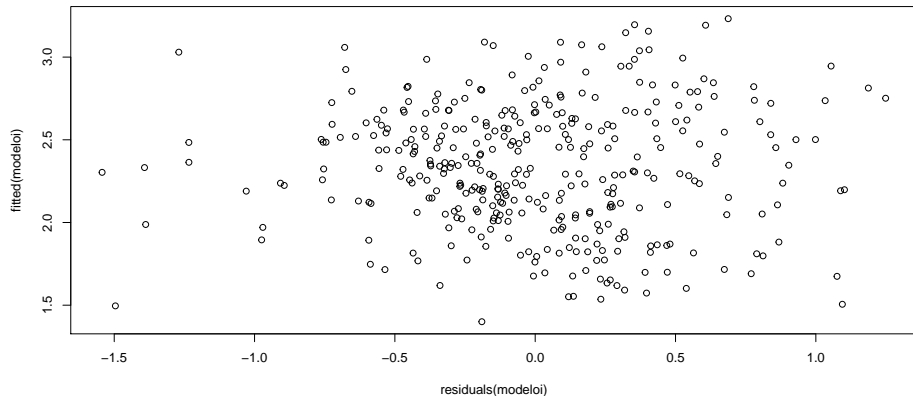
Heterocedasticidade: gpa3

```
library(sandwich)
library(lmtest)
# Versão robusta (à Heterocedasticidade) do teste F
waldtest(modelor, modeloi, vcov = vcovHC(modeloi, type = "HC"))

## Wald test
##
## Model 1: cumgpa ~ sat + hsperc + tothrs + female
## Model 2: cumgpa ~ sat + hsperc + tothrs + female + black + whi
##   Res.Df Df       F Pr(>F)
## 1      361
## 2      359  2 0.7478 0.4741
```

Heterocedasticidade: gpa3

```
plot(residuals(modeloi),fitted(modeloi))
```



Heterocedasticidade: crime1

$$\text{narr86} = \beta_0 + \beta_1 \text{pcnv} + \beta_2 \text{avgsen} + \beta_3 \text{avgsen}^2 + \beta_4 \text{ptime86} + \beta_5 \text{qemp86} + \beta_6 \text{black} + \beta_7 \text{hispan} + u$$

Heterocedasticidade: crime1

$$narr86 = \beta_0 + \beta_1 pcnv + \beta_2 avg\textit{sen} + \beta_3 avg\textit{sen}^2 + \beta_4 ptime86 + \beta_5 qemp86 + \beta_6 black + \beta_7 hispan + u$$

- ▶ Fazer o *teste t* para testar a significância das variáveis
- ▶ Fazer o *teste F* para testar

$$H_0 : \beta_{avg\textit{sen}} = 0 \quad \beta_{avg\textit{sen}^2} = 0 \quad \text{vs} \quad H_1 : H_0 \text{ não é verdade}$$

Heterocedasticidade: crime1

$$narr86 = \beta_0 + \beta_1 pcnv + \beta_2 avgsen + \beta_3 avgsen^2 + \beta_4 ptime86 + \beta_5 qemp86 + \beta_6 black + \beta_7 hispan + u$$

- ▶ Fazer o *teste t* para testar a significância das variáveis
- ▶ Fazer o *teste F* para testar

$$H_0 : \beta_{avgsen} = 0 \quad \beta_{avgsen^2} = 0 \quad \text{vs} \quad H_1 : H_0 \text{ não é verdade}$$

```
modeloi = lm(narr86~pcnv+avgsen+I(avgsen^2)+ptime86+
             qemp86+inc86+black+hispan, data = crime1)
modelor = lm(narr86~pcnv+ptime86+qemp86+inc86+black+
             hispan, data = crime1)
```

Heterocedasticidade: crime1

Comparação

##	betas	Std	White	P-valor	P-valor_W
## (Intercept)	0.56701	0.03606	0.04021	0.00000	0.00000
## pcnv	-0.13560	0.04037	0.03357	0.00079	0.00006
## avgsen	0.01784	0.00970	0.01011	0.06587	0.07763
## I(avgsen^2)	-0.00052	0.00030	0.00021	0.08226	0.01283
## ptime86	-0.03936	0.00869	0.00621	0.00001	0.00000
## qemp86	-0.05051	0.01443	0.01418	0.00047	0.00037
## inc86	-0.00148	0.00034	0.00023	0.00001	0.00000
## black	0.32460	0.04542	0.05842	0.00000	0.00000
## hispan	0.19338	0.03970	0.04023	0.00000	0.00000

Heterocedasticidade: crime1

$$H_0 : \beta_{avgsen} = 0 \quad \beta_{avgsen^2} = 0 \quad \text{vs} \quad H_1 : H_0 \text{ não é verdade}$$

Teste F clássico

```
anova(modelor,modeloi)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: narr86 ~ pcnv + ptime86 + qemp86 + inc86 + black + h
```

```
## Model 2: narr86 ~ pcnv + avgsen + I(avgsen^2) + ptime86 + qemp
```

```
##      black + hispan
```

```
##      Res.Df      RSS Df Sum of Sq      F Pr(>F)
```

```
## 1      2718 1866.4
```

```
## 2      2716 1864.0  2      2.3716 1.7278 0.1779
```


Heterocedasticidade: crime1

$$H_0 : \beta_{avgsen} = 0 \quad \beta_{avgsen^2} = 0 \quad \text{vs} \quad H_1 : H_0 \text{ não é verdade}$$

```
# Teste F robusto
```

```
waldtest(modelor, modeloi, vcov = vcovHC(modeloi, type = "HC"))
```

```
## Wald test
```

```
##
```

```
## Model 1: narr86 ~ pcnv + ptime86 + qemp86 + inc86 + black + hispan
```

```
## Model 2: narr86 ~ pcnv + avgsen + I(avgsen^2) + ptime86 + qemp86 +
```

```
##      black + hispan
```

```
##      Res.Df Df      F    Pr(>F)
```

```
## 1      2718
```

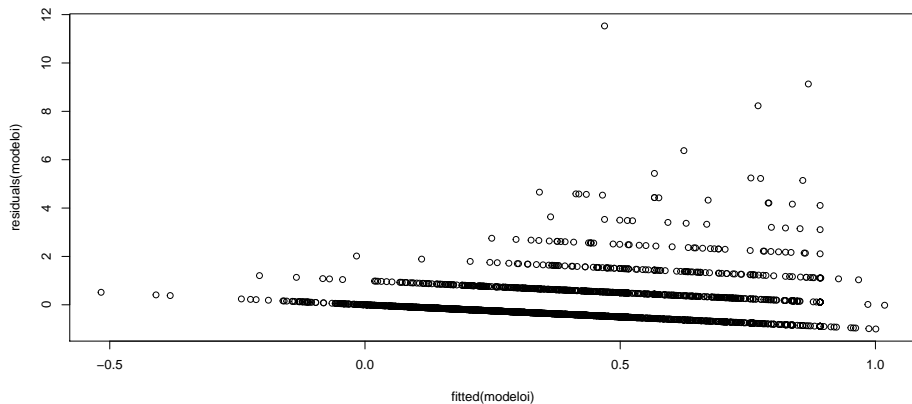
```
## 2      2716    2 4.7806 0.008461 **
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Heterocedasticidade: crime1

```
plot(fitted(modeloi), residuals(modeloi))
```



Heterocedasticidade: Teste para Heterocedasticidade

- ▶ Na presença de heterocedasticidade, embora os estimadores MQO continuem sendo não viesados, eles não são mais o melhor estimador linear não viesado.

Heterocedasticidade: Teste para Heterocedasticidade

- ▶ Na presença de heterocedasticidade, embora os estimadores MQO continuem sendo não viesados, eles não são mais o melhor estimador linear não viesado.
- ▶ Embora o estimador robusto de White funcione bem em casos heterocedásticos e homocedásticos, sob HRLM6 (Normalidade) as estatísticas t e F são exatas e não aproximadas como com o estimador de White.

Heterocedasticidade: Teste para Heterocedasticidade

- ▶ Na presença de heterocedasticidade, embora os estimadores MQO continuem sendo não viesados, eles não são mais o melhor estimador linear não viesado.
- ▶ Embora o estimador robusto de White funcione bem em casos heterocedásticos e homocedásticos, sob HRLM6 (Normalidade) as estatísticas t e F são exatas e não aproximadas como com o estimador de White.
- ▶ Às vezes, graficamente não é tão fácil identificar a presença de heterocedasticidade, mas podemos utilizar um teste para testar

$$H_0 : \underbrace{V(u|X) = E(u^2|X) = E(u^2)}_{\text{HRLM4: } E(u|X)=0} = \sigma^2 I$$

Heterocedasticidade: Teste para Heterocedasticidade

Sob H_0 (Homocedasticidade) não deve existir nenhuma relação entre u^2 com alguma variável explicativa

Teste Breusch-Pagan da Heterocedasticidade

$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \dots + \delta_k x_k + \nu$$

Teste de White para a Heterocedasticidade

$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \dots + \delta_k x_k + \delta_{k+1} x_1^2 + \dots + \delta_{2k} x_k^2 + \delta_{2k+1} x_1 x_2 + \delta_{2k+2} x_1 x_3 + \dots + \nu$$

H_0 implica que os coeficientes δ sejam todos iguais a zero (ou seja, homocedasticidade).

Heterocedasticidade: Teste para Heterocedasticidade

```
modelo = lm(price~lotsize+sqrft + bdrms, data = hprice1)
round(summary(modelo)$coef,4)
```

##	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	-21.7703	29.4750	-0.7386	0.4622
## lotsize	0.0021	0.0006	3.2201	0.0018
## sqrft	0.1228	0.0132	9.2751	0.0000
## bdrms	13.8525	9.0101	1.5374	0.1279

Heterocedasticidade: Teste para Heterocedasticidade

```
# Test Breusch-Pagan
```

```
bptest(modelo, studentize = TRUE)
```

```
##
```

```
## studentized Breusch-Pagan test
```

```
##
```

```
## data:  modelo
```

```
## BP = 14.092, df = 3, p-value = 0.002782
```


Heterocedasticidade: Teste para Heterocedasticidade

```
# Teste de White
```

```
bptest(modelo, ~ lotsize*sqrft + lotsize*bdrms + sqrft*bdrms + I(  
    I(sqrft^2) + I(bdrms^2), studentize = TRUE, data = hprice)
```

```
##
```

```
## studentized Breusch-Pagan test
```

```
##
```

```
## data: modelo
```

```
## BP = 33.732, df = 9, p-value = 9.953e-05
```

Heterocedasticidade: Teste para Heterocedasticidade

```
modelo = lm(log(price)~log(lotsize)+log(sqrft) + bdrms, data = hp)  
round(summary(modelo)$coef,4)
```

##	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	-1.2970	0.6513	-1.9915	0.0497
## log(lotsize)	0.1680	0.0383	4.3877	0.0000
## log(sqrft)	0.7002	0.0929	7.5403	0.0000
## bdrms	0.0370	0.0275	1.3424	0.1831

Heterocedasticidade: Teste para Heterocedasticidade

```
# Test Breusch-Pagan
```

```
bptest(modelo, studentize = TRUE)
```

```
##
```

```
## studentized Breusch-Pagan test
```

```
##
```

```
## data:  modelo
```

```
## BP = 4.2232, df = 3, p-value = 0.2383
```

Heterocedasticidade: Teste para Heterocedasticidade

```
# Teste de White
```

```
bptest(modelo, ~ log(lotsize)*log(sqrft) +  
          log(lotsize)*bdrms + log(sqrft)*bdrms +  
          I(log(lotsize)^2) + I(log(sqrft)^2) + I(bdrms^2),  
          studentize = TRUE, data = hprice1)
```

```
##
```

```
## studentized Breusch-Pagan test
```

```
##
```

```
## data:  modelo
```

```
## BP = 9.5495, df = 9, p-value = 0.3882
```

Heterocedasticidade

- ▶ Os testes de heterocedasticidade BP e White são construídos sob HRML1–HRML4, se uma das hipóteses não for verificada afeta os testes.
- ▶ Algumas vezes $E(y|X)$ está mal-especificada, nesse caso os testes de heterocedasticidade podem rejeitar H_0 mesmo quando $V(u|X) = \sigma^2 I$
- ▶ Devemos primeiro verificar a má-especificação antes de testarmos heterocedasticidade
- ▶ **Outra alternativa:** Mínimos Quadrados Ponderados (MQP)

Leituras recomendadas

Leituras recomendadas

- ▶ Montgomery, D. C; Peck, E. A; Vining, G. *Introduction to Linear Regression Analysis*. (2012). Wiley, 5ed. – **Cap 4**
- ▶ Wooldridge, Jeffrey M. *Introdução à Econometria: Uma abordagem moderna*. (2016). Cengage Learning. – **Cap 8**