

ANÁLISE DE DADOS MULTIVARIADOS I - REGRESSÃO

(AULA 03)

Novembro e dezembro de 2018

Reinaldo Soares de Camargo

Modelos de Regressão

- Regressão Linear Múltipla

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \epsilon_i$$

- Análise de regressão múltipla possibilita *ceteris paribus* (outros fatores constantes), pois permite controlar muitos outros fatores que afetam a variável dependente simultaneamente.
- Isso auxilia no teste de teorias econômicas e na avaliação de impactos de políticas públicas, quando possuímos dados não-experimentais.
- Ao utilizar mais fatores na explicação de y , uma maior variação de y será explicada pelo modelo.
- Este é o modelo mais utilizado nas ciências sociais.
- O método de MQO é usado para estimar os parâmetros do modelo de regressão múltipla.

Modelos de Regressão

- Modelos com k variáveis independentes

$$y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_k X_k + \epsilon_i$$

- Esse é o modelo de regressão linear múltipla geral ou, simplesmente, modelo de regressão múltipla.
- Há k+ 1 parâmetros populacionais desconhecidos, já que temos k variáveis independentes e um intercepto.
- Os parâmetros β_1 a β_k são chamados de parâmetros de inclinação, mesmo que eles não tenham exatamente este significado.
- A regressão é “linear” porque é linear nos β_j , mesmo que seja uma relação não-linear entre a variável dependente e as variáveis independentes:

$$y = \beta_0 + \beta_1 \log(x_1) + \beta_2 x_2 + \beta_3 x_2^2 + u$$

Modelos de Regressão

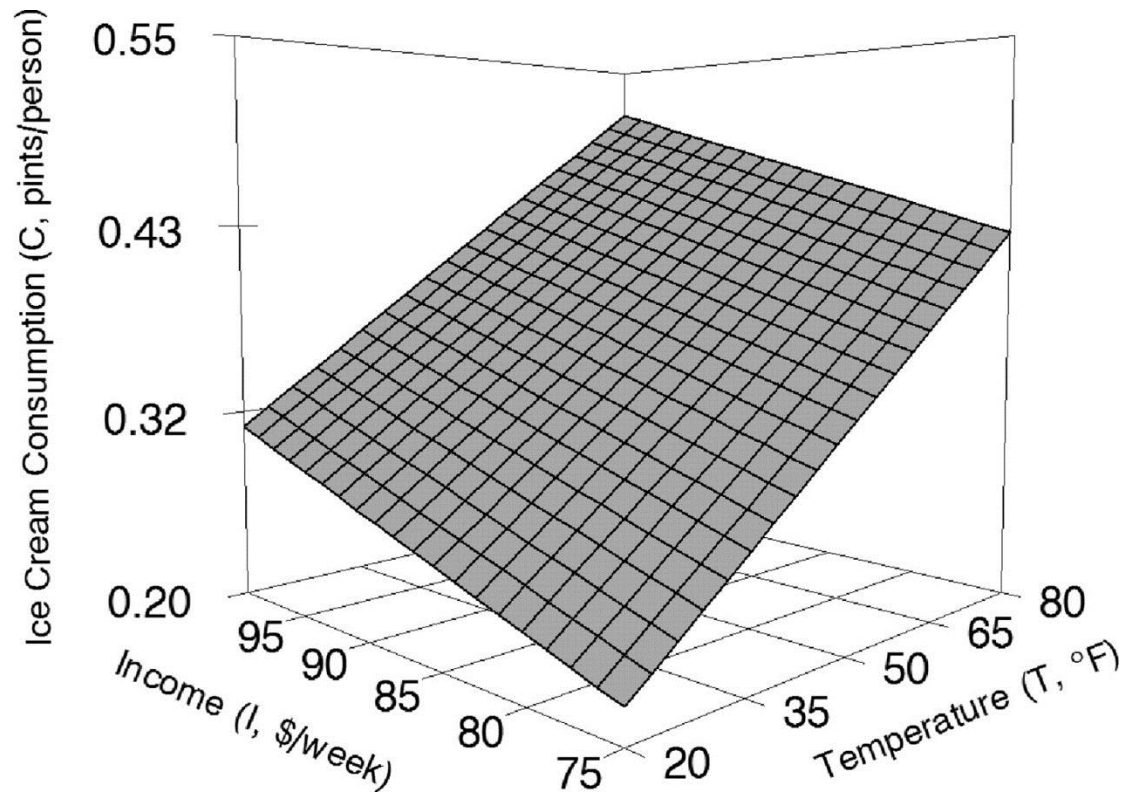
- Modelos com duas variáveis independentes

$$y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon_i$$

- β_0 é o intercepto é o valor previsto de Y quando todas as variáveis independentes são iguais a zero.
- β_1 mede a variação em y com relação a x1, mantendo os outros fatores constantes.
- β_2 mede a variação em y com relação a x2, mantendo os outros fatores constantes.

Modelos de Regressão

- Regressão linear múltipla (mais de uma variável explicativa)



Modelos de Regressão

- Pressupostos do modelo de regressão múltipla

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \epsilon_i$$

OU

$$Y = X\beta + \epsilon$$

1. O termo de erro possui média zero: $E(\epsilon_i) = 0$
2. Os erros são normalmente distribuídos.
3. A variável x não é aleatória e deve assumir pelo menos dois valores diferentes.
4. Variância do erro é constante: $\text{var}(\epsilon_i) = \sigma^2$
5. Os erros não são autocorrelacionados : $E(\epsilon_i \epsilon_j) = 0, i \neq j$
6. Cada variável independente X_i não pode ser combinação linear das demais

Modelos de Regressão

- **Aplicação em R**

- Com base nos dados IDH_Brasil_2010.csv
- Estime um modelo de regressão linear múltipla para explicar renda percapita em função e idhm educação.

```
> mod1 <- lm(renda_per_capita ~ esperanza_vida_ao_nascer + IDHM_educacao  
+ , data = dados)  
> summary(mod1)
```

- Espoço o gráfico dos resíduos do modelo estimado

```
> par(mfrow=c(2,2))  
> plot(mod1)
```

- Interprete os resultados.

Modelos de Regressão

```
> mod1 <- lm(renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao,  
+ data = dados)  
> summary(mod1)
```

Call:
lm(formula = renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao,
data = dados)

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|-------|---------|
| -358.85 | -82.99 | -11.48 | 63.59 | 1113.90 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------------|------------|------------|---------|----------|
| (Intercept) | -3182.2559 | 56.1557 | -56.67 | <2e-16 |
| esperanca_vida_ao_nascer | 40.8346 | 0.8935 | 45.70 | <2e-16 |
| IDHM_educacao | 1236.5080 | 25.6647 | 48.18 | <2e-16 |

Residual standard error: 126.8 on 5561 degrees of freedom
Multiple R-squared: 0.7283, Adjusted R-squared: 0.7282
F-statistic: 7454 on 2 and 5561 DF, p-value: < 2.2e-16

Variável dependente no modelo de regressão

Modelos de Regressão

```
> mod1 <- lm(renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao  
+ , data = dados)  
> summary(mod1)
```

Call:
lm(formula = renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao,
data = dados)

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|-------|---------|
| -358.85 | -82.99 | -11.48 | 63.59 | 1113.90 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------------|------------|------------|---------|----------|
| (Intercept) | -3182.2559 | 56.1557 | -56.67 | <2e-16 |
| esperanca_vida_ao_nascer | 40.8346 | 0.8935 | 45.70 | <2e-16 |
| IDHM_educacao | 1236.5080 | 25.6647 | 48.18 | <2e-16 |

Residual standard error: 126.8 on 5561 degrees of freedom
Multiple R-squared: 0.7283, Adjusted R-squared: 0.7282
F-statistic: 7454 on 2 and 5561 DF, p-value: < 2.2e-16

Variáveis independentes no modelo de regressão

Modelos de Regressão

```
> mod1 <- lm(renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao  
+ , data = dados)  
> summary(mod1)
```

Call:
lm(formula = renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao,
data = dados)

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|-------|---------|
| -358.85 | -82.99 | -11.48 | 63.59 | 1113.90 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------------|------------|------------|---------|----------|
| (Intercept) | -3182.2559 | 56.1557 | -56.67 | <2e-16 |
| esperanca_vida_ao_nascer | 40.8346 | 0.8935 | 45.70 | <2e-16 |
| IDHM_educacao | 1236.5080 | 25.6647 | 48.18 | <2e-16 |

Residual standard error: 126.8 on 5561 degrees of freedom
Multiple R-squared: 0.7283, Adjusted R-squared: 0.7282
F-statistic: 7454 on 2 and 5561 DF, p-value: < 2.2e-16

Base de dados que deve conter as variáveis do modelo

Modelos de Regressão

```
> mod1 <- lm(renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao  
+ , data = dados)  
> summary(mod1)
```

Call:
lm(formula = renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao,
data = dados)

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|-------|---------|
| -358.85 | -82.99 | -11.48 | 63.59 | 1113.90 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------------|------------|------------|---------|----------|
| (Intercept) | -3182.2559 | 56.1557 | -56.67 | <2e-16 |
| esperanca_vida_ao_nascer | 40.8346 | 0.8935 | 45.70 | <2e-16 |
| IDHM_educacao | 1236.5080 | 25.6647 | 48.18 | <2e-16 |

Residual standard error: 126.8 on 5561 degrees of freedom
Multiple R-squared: 0.7283, Adjusted R-squared: 0.7282
F-statistic: 7454 on 2 and 5561 DF, p-value: < 2.2e-16

Comando utilizado para mostrar os resultados do modelo

Modelos de Regressão

```
> mod1 <- lm(renda_per_capita ~ esperanza_vida_ao_nascer + IDHM_educacao  
+ , data = dados)  
> summary(mod1)
```

```
Call:  
lm(formula = renda_per_capita ~ esperanza_vida_ao_nascer + IDHM_educacao,  
    data = dados)
```

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|-------|---------|
| -358.85 | -82.99 | -11.48 | 63.59 | 1113.90 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------------|------------|------------|---------|----------|
| (Intercept) | -3182.2559 | 56.1557 | -56.67 | <2e-16 |
| esperanca_vida_ao_nascer | 40.8346 | 0.8935 | 45.70 | <2e-16 |
| IDHM_educacao | 1236.5080 | 25.6647 | 48.18 | <2e-16 |

Residual standard error: 126.8 on 5561 degrees of freedom
Multiple R-squared: 0.7283, Adjusted R-squared: 0.7282
F-statistic: 7454 on 2 and 5561 DF, p-value: < 2.2e-16

Estatísticas dos resíduos do modelo (do termo ε)

Modelos de Regressão

```
> mod1 <- lm(renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao  
+ , data = dados)  
> summary(mod1)
```

Call:
lm(formula = renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao,
data = dados)

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|-------|---------|
| -358.85 | -82.99 | -11.48 | 63.59 | 1113.90 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------------|------------|------------|---------|----------|
| (Intercept) | -3182.2559 | 56.1557 | -56.67 | <2e-16 |
| esperanca_vida_ao_nascer | 40.8346 | 0.8935 | 45.70 | <2e-16 |
| IDHM_educacao | 1236.5080 | 25.6647 | 48.18 | <2e-16 |

Residual standard error: 126.8 on 5561 degrees of freedom
Multiple R-squared: 0.7283, Adjusted R-squared: 0.7282
F-statistic: 7454 on 2 and 5561 DF, p-value: < 2.2e-16

Coeficientes estimados do modelo: $\beta_0, \beta_1, \beta_2, \dots, \beta_k$; estatísticas de teste dos coeficientes estimados.

Modelos de Regressão

```
> mod1 <- lm(renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao  
+ , data = dados)  
> summary(mod1)
```

Call:
lm(formula = renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao,
data = dados)

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|-------|---------|
| -358.85 | -82.99 | -11.48 | 63.59 | 1113.90 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------------|------------|------------|---------|----------|
| (Intercept) | -3182.2559 | 56.1557 | -56.67 | <2e-16 |
| esperanca_vida_ao_nascer | 40.8346 | 0.8935 | 45.70 | <2e-16 |
| IDHM_educacao | 1236.5080 | 25.6647 | 48.18 | <2e-16 |

Residual standard error: 126.8 on 5561 degrees of freedom
Multiple R-squared: 0.7283, Adjusted R-squared: 0.7282
F-statistic: 7454 on 2 and 5561 DF, p-value: < 2.2e-16

Estimativas dos coeficientes β_k . Representa a variação na variável Y explicada pela variação de cada variável independente *ceteris paribus*

Modelos de Regressão

```
> mod1 <- lm(renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao  
+ , data = dados)  
> summary(mod1)
```

Call:
lm(formula = renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao,
data = dados)

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|-------|---------|
| -358.85 | -82.99 | -11.48 | 63.59 | 1113.90 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------------|------------|------------|---------|----------|
| (Intercept) | -3182.2559 | 56.1557 | -56.67 | <2e-16 |
| esperanca_vida_ao_nascer | 40.8346 | 0.8935 | 45.70 | <2e-16 |
| IDHM_educacao | 1236.5080 | 25.6647 | 48.18 | <2e-16 |

Residual standard error: 126.8 on 5561 degrees of freedom
Multiple R-squared: 0.7283, Adjusted R-squared: 0.7282
F-statistic: 7454 on 2 and 5561 DF, p-value: < 2.2e-16

β_0 (Intercept) valor previsto de renda_per_capita quando esperança de vida ao nascer e Idhm_educação forem iguais a zero.

Modelos de Regressão

```
> mod1 <- lm(renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao  
+ , data = dados)  
> summary(mod1)
```

Call:
lm(formula = renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao,
data = dados)

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|-------|---------|
| -358.85 | -82.99 | -11.48 | 63.59 | 1113.90 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------------|------------|------------|---------|----------|
| (Intercept) | -3182.2559 | 56.1557 | -56.67 | <2e-16 |
| esperanca_vida_ao_nascer | 10.8346 | 0.8935 | 45.70 | <2e-16 |
| IDHM_educacao | 1236.5080 | 25.6647 | 48.18 | <2e-16 |

Residual standard error: 126.8 on 5561 degrees of freedom
Multiple R-squared: 0.7283, Adjusted R-squared: 0.7282
F-statistic: 7454 on 2 and 5561 DF, p-value: < 2.2e-16

Interpretação: Para localidades com esperança de vida ao nascer e idhm_educação iguais a zero, a renda per capita estimada é de –R\$ 3.182,25.

Modelos de Regressão

```
> mod1 <- lm(renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao  
+ , data = dados)  
> summary(mod1)
```

Call:
lm(formula = renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao,
data = dados)

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|-------|---------|
| -358.85 | -82.99 | -11.48 | 63.59 | 1113.90 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------------|------------|------------|---------|----------|
| (Intercept) | -2102.2559 | 56.1557 | -56.67 | <2e-16 |
| esperanca_vida_ao_nascer | 40.8346 | 0.8935 | 45.70 | <2e-16 |
| IDHM_educacao | 1238.5080 | 25.6647 | 48.18 | <2e-16 |

Residual standard error: 126.8 on 5561 degrees of freedom
Multiple R-squared: 0.7283, Adjusted R-squared: 0.7282
F-statistic: 7454 on 2 and 5561 DF, p-value: < 2.2e-16

β_1 mede o efeito do aumento de uma unidade na esperança de vida ao nascer na renda per capita, mantendo IDHM educação constantes (ceteris paribus)

Modelos de Regressão

```
> mod1 <- lm(renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao  
+ , data = dados)  
> summary(mod1)
```

Call:
lm(formula = renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao,
data = dados)

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|-------|---------|
| -358.85 | -82.99 | -11.48 | 63.59 | 1113.90 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------------|------------|------------|---------|----------|
| (Intercept) | -2102.2559 | 56.1557 | -56.67 | <2e-16 |
| esperanca_vida_ao_nascer | 40.8346 | 0.8935 | 45.70 | <2e-16 |
| IDHM_educacao | 1238.5080 | 25.6647 | 48.18 | <2e-16 |

Residual standard error: 126.8 on 5561 degrees of freedom
Multiple R-squared: 0.7283, Adjusted R-squared: 0.7282
F-statistic: 7454 on 2 and 5561 DF, p-value: < 2.2e-16

Interpretação: O aumento de uma unidade na esperança de vida ao nascer (de um ano), aumenta a renda per capita da localidade em R\$ 40,83 , mantedo constante idhm_educação.

Modelos de Regressão

```
> mod1 <- lm(renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao  
+ , data = dados)  
> summary(mod1)
```

Call:
lm(formula = renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao,
data = dados)

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|-------|---------|
| -358.85 | -82.99 | -11.48 | 63.59 | 1113.90 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------------|------------|------------|---------|----------|
| (Intercept) | -3182.2559 | 56.1557 | -56.67 | <2e-16 |
| esperanca_vida_ao_nascer | 10.8246 | 0.8935 | 45.70 | <2e-16 |
| IDHM_educacao | 1236.5080 | 25.6647 | 48.18 | <2e-16 |

Residual standard error: 126.8 on 5561 degrees of freedom
Multiple R-squared: 0.7283, Adjusted R-squared: 0.7282
F-statistic: 7454 on 2 and 5561 DF, p-value: < 2.2e-16

β_1 mede o efeito do aumento de uma unidade no IDHM_educação na renda per capita, mantendo esperança de vida ao nascer constante(ceteris paribus)

Modelos de Regressão

```
> mod1 <- lm(renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao  
+ , data = dados)  
> summary(mod1)
```

Call:
lm(formula = renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao,
data = dados)

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|-------|---------|
| -358.85 | -82.99 | -11.48 | 63.59 | 1113.90 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------------|------------|------------|---------|----------|
| (Intercept) | -3182.2559 | 56.1557 | -56.67 | <2e-16 |
| esperanca_vida_ao_nascer | 10.8246 | 0.8935 | 45.70 | <2e-16 |
| IDHM_educacao | 1236.5080 | 25.6647 | 48.18 | <2e-16 |

Residual standard error: 126.8 on 5561 degrees of freedom
Multiple R-squared: 0.7283, Adjusted R-squared: 0.7282
F-statistic: 7454 on 2 and 5561 DF, p-value: < 2.2e-16

Interpretação: O aumento de uma unidade no idhm_educação aumenta a renda per capita da localidade em R\$ 1.236,50 , mantedo constante a esperança de vida ao nascer.

Modelos de Regressão

```
> mod1 <- lm(renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao  
+ , data = dados)  
> summary(mod1)
```

Call:
lm(formula = renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao,
data = dados)

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|-------|---------|
| -358.85 | -82.99 | -11.48 | 63.59 | 1113.90 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------------|------------|------------|---------|----------|
| (Intercept) | -3182.2559 | 56.1557 | -56.67 | <2e-16 |
| esperanca_vida_ao_nascer | 40.8345 | 0.8935 | 45.70 | <2e-16 |
| IDHM_educacao | 1236.5080 | 25.6647 | 48.18 | <2e-16 |

Residual standard error: 126.8 on 5561 degrees of freedom
Multiple R-squared: 0.7283, Adjusted R-squared: 0.7282
F-statistic: 7454 on 2 and 5561 DF, p-value: < 2.2e-16

Erros padrões das estimativas dos coeficientes (β) servem para construir intervalos de confiança para as estimativas, e calcular o t-value.

Modelos de Regressão

```
> mod1 <- lm(renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao  
+ , data = dados)  
> summary(mod1)
```

Call:
lm(formula = renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao,
data = dados)

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|-------|---------|
| -358.85 | -82.99 | -11.48 | 63.59 | 1113.90 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------------|------------|------------|---------|----------|
| (Intercept) | -3182.2559 | 56.1557 | -56.67 | <2e-16 |
| esperanca_vida_ao_nascer | 40.8346 | 0.8935 | 45.70 | <2e-16 |
| IDHM_educacao | 1236.5080 | 25.6647 | 48.18 | <2e-16 |

Residual standard error: 126.8 on 5561 degrees of freedom
Multiple R-squared: 0.7283, Adjusted R-squared: 0.7282
F-statistic: 7454 on 2 and 5561 DF, p-value: < 2.2e-16

Valor da estatística t – Student das estimativas dos coeficientes (β) , serve para realização de testes de hipótese e cálculo do p-value ($\Pr(>|t|)$)

Modelos de Regressão

```
> mod1 <- lm(renda_per_capita ~ esperanza_vida_ao_nascer + IDHM_educacao  
+ , data = dados)  
> summary(mod1)
```

Call:
lm(formula = renda_per_capita ~ esperanza_vida_ao_nascer + IDHM_educacao,
data = dados)

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|-------|---------|
| -358.85 | -82.99 | -11.48 | 63.59 | 1113.90 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------------|------------|------------|---------|----------|
| (Intercept) | -3182.2559 | 56.1557 | -56.67 | <2e-16 |
| esperanca_vida_ao_nascer | 40.8346 | 0.8935 | 45.70 | <2e-16 |
| IDHM_educacao | 1236.5080 | 25.6647 | 48.18 | <2e-16 |

Residual standard error: 126.8 on 5561 degrees of freedom

Multiple R-squared: 0.7283, Adjusted R-squared: 0.7282

F-statistic: 7454 on 2 and 5561 DF, p-value: < 2.2e-16

P-value das estimativas dos coeficientes (β_k). Serve para testar a significância individual de cada variável independente. É o menor nível de significância utilizado para rejeitar a hipótese nula: $H_0: \hat{\beta}_k = 0$ a um dado nível de significância (5%, 10%).

Modelos de Regressão

```
> mod1 <- lm(renda_per_capita ~ esperanza_vida_ao_nascer + IDHM_educacao  
+ , data = dados)  
> summary(mod1)
```

```
Call:  
lm(formula = renda_per_capita ~ esperanza_vida_ao_nascer + IDHM_educacao,  
    data = dados)
```

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|-------|---------|
| -358.85 | -82.99 | -11.48 | 63.59 | 1113.90 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------------|------------|------------|---------|----------|
| (Intercept) | -3182.2559 | 56.1557 | -56.67 | <2e-16 |
| esperanca_vida_ao_nascer | 40.8346 | 0.8935 | 45.70 | <2e-16 |
| IDHM_educacao | 1236.5080 | 25.6647 | 48.18 | <2e-16 |

Residual standard error: 126.8 on 5561 degrees of freedom

Multiple R-squared: 0.7283, Adjusted R-squared: 0.7282

F-statistic: 7454 on 2 and 5561 DF, p-value: < 2.2e-16

Desvio padrão dos resíduos, utilizado para calcular o R^2 e as estatísticas F para o teste de significância conjunta do modelo.

Modelos de Regressão

```
> mod1 <- lm(renda_per_capita ~ esperanza_vida_ao_nascer + IDHM_educacao  
+ , data = dados)  
> summary(mod1)
```

Call:
lm(formula = renda_per_capita ~ esperanza_vida_ao_nascer + IDHM_educacao,
data = dados)

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|-------|---------|
| -358.85 | -82.99 | -11.48 | 63.59 | 1113.90 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------------|------------|------------|---------|----------|
| (Intercept) | -3182.2559 | 56.1557 | -56.67 | <2e-16 |
| esperanca_vida_ao_nascer | 40.8346 | 0.8935 | 45.70 | <2e-16 |
| IDHM_educacao | 1236.5080 | 25.6647 | 48.18 | <2e-16 |

Residual standard error: 126.8 on 5561 degrees of freedom

Multiple R-squared: 0.7283. Adjusted R-squared: 0.7282

F-statistic: 7434 on 2 and 5561 DF, p-value: < 2.2e-16

R^2 mede o grau de ajuste geral do modelo, assim como um teste para indicar o quanto o conjunto das variáveis do modelo explica variações em y .

Modelos de Regressão

```
> mod1 <- lm(renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao  
+ , data = dados)  
> summary(mod1)
```

```
Call:  
lm(formula = renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao,  
    data = dados)
```

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|-------|---------|
| -358.85 | -82.99 | -11.48 | 63.59 | 1113.90 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------------|------------|------------|---------|----------|
| (Intercept) | -3182.2559 | 56.1557 | -56.67 | <2e-16 |
| esperanca_vida_ao_nascer | 40.8346 | 0.8935 | 45.70 | <2e-16 |
| IDHM_educacao | 1236.5080 | 25.6647 | 48.18 | <2e-16 |

Residual standard error: 126.8 on 5561 degrees of freedom

Multiple R-squared: 0.7283. Adjusted R-squared: 0.7282

F-statistic: 7434 on 2 and 5561 DF, p-value: < 2.2e-16

Interpretação: 72,83% das das variações da renda per capita são explicadas por esperança de vida ao nascer e idhm educação.

Modelos de Regressão

```
> mod1 <- lm(renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao  
+ , data = dados)  
> summary(mod1)
```

Call:
lm(formula = renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao,
data = dados)

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|-------|---------|
| -358.85 | -82.99 | -11.48 | 63.59 | 1113.90 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------------|------------|------------|---------|----------|
| (Intercept) | -3182.2559 | 56.1557 | -56.67 | <2e-16 |
| esperanca_vida_ao_nascer | 40.8346 | 0.8935 | 45.70 | <2e-16 |
| IDHM_educacao | 1236.5080 | 25.6647 | 48.18 | <2e-16 |

Residual standard error: 126.8 on 5561 degrees of freedom

Multiple R-squared: 0.7283, Adjusted R-squared: 0.7282

F-statistic: 7454 on 2 and 5561 DF, p-value: < 2.2e-16

Como a inclusão de variáveis na equação, independente de fazerem sentido ou não, o R^2 sempre aumenta, uma alternativa é utilizar o R^2 ajustado para verificar o ajuste de modelos de regressão múltipla. Para seleção de modelos utilizar o de maior R^2_a .

Modelos de Regressão

```
> mod1 <- lm(renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao  
+ , data = dados)  
> summary(mod1)
```

```
Call:  
lm(formula = renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao,  
    data = dados)
```

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|-------|---------|
| -358.85 | -82.99 | -11.48 | 63.59 | 1113.90 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------------|------------|------------|---------|----------|
| (Intercept) | -3182.2559 | 56.1557 | -56.67 | <2e-16 |
| esperanca_vida_ao_nascer | 40.8346 | 0.8935 | 45.70 | <2e-16 |
| IDHM_educacao | 1236.5080 | 25.6647 | 48.18 | <2e-16 |

Residual standard error: 126.8 on 5561 degrees of freedom

Multiple R-squared: 0.7283, Adjusted R-squared: 0.7282

F-statistic: 7454 on 2 and 5561 DF, p-value: < 2.2e-16

Interpretação: 72,82% das variações da renda percapita são explicadas por variações nas variáveis esperança de vida ao nascer e idhm educação.

Modelos de Regressão

```
> mod1 <- lm(renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao  
+ , data = dados)  
> summary(mod1)
```

```
Call:  
lm(formula = renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao,  
    data = dados)
```

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|-------|---------|
| -358.85 | -82.99 | -11.48 | 63.59 | 1113.90 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------------|------------|------------|---------|----------|
| (Intercept) | -3182.2559 | 56.1557 | -56.67 | <2e-16 |
| esperanca_vida_ao_nascer | 40.8346 | 0.8935 | 45.70 | <2e-16 |
| IDHM_educacao | 1236.5080 | 25.6647 | 48.18 | <2e-16 |

Residual standard error: 126.8 on 5561 degrees of freedom

Multiple R-squared: 0.7283, Adjusted R-squared: 0.7282

F-statistic: 7454 on 2 and 5561 DF, p-value: < 2.2e-16

Estatística F, utilizada para teste de significância conjunta do modelo estimado. P-value, menor nível de significância utilizado para rejeitar a hipótese nula: $H_0 : \beta_1 = \beta_2 = 0$;

Modelos de Regressão

```
> mod1 <- lm(renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao  
+ , data = dados)  
> summary(mod1)
```

```
Call:  
lm(formula = renda_per_capita ~ esperanca_vida_ao_nascer + IDHM_educacao,  
    data = dados)
```

Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|-------|---------|
| -358.85 | -82.99 | -11.48 | 63.59 | 1113.90 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------------|------------|------------|---------|----------|
| (Intercept) | -3182.2559 | 56.1557 | -56.67 | <2e-16 |
| esperanca_vida_ao_nascer | 40.8346 | 0.8935 | 45.70 | <2e-16 |
| IDHM_educacao | 1236.5080 | 25.6647 | 48.18 | <2e-16 |

Residual standard error: 126.8 on 5561 degrees of freedom

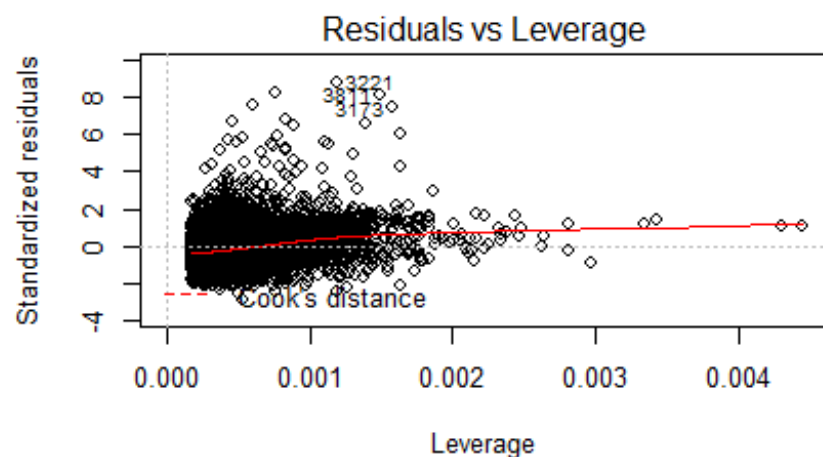
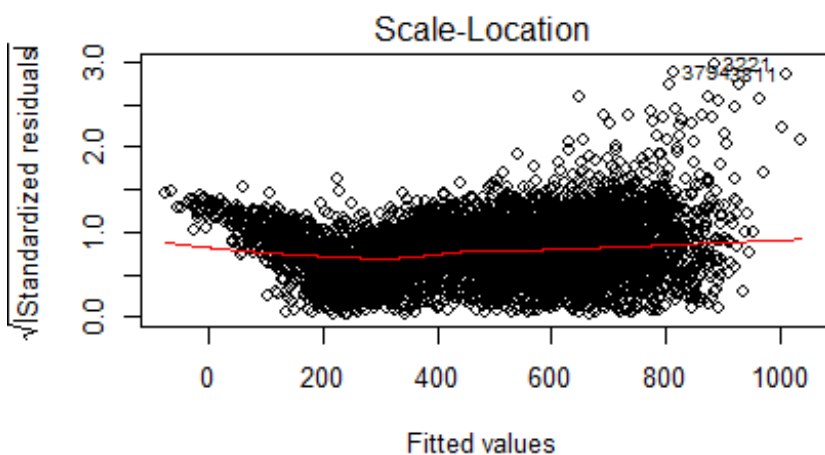
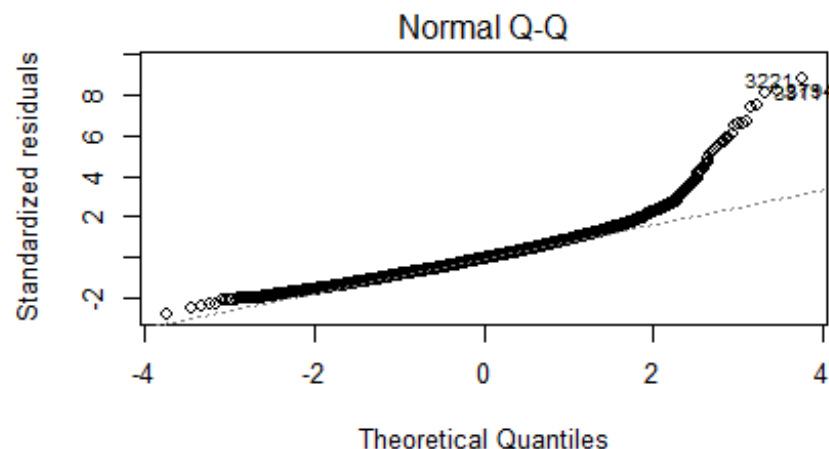
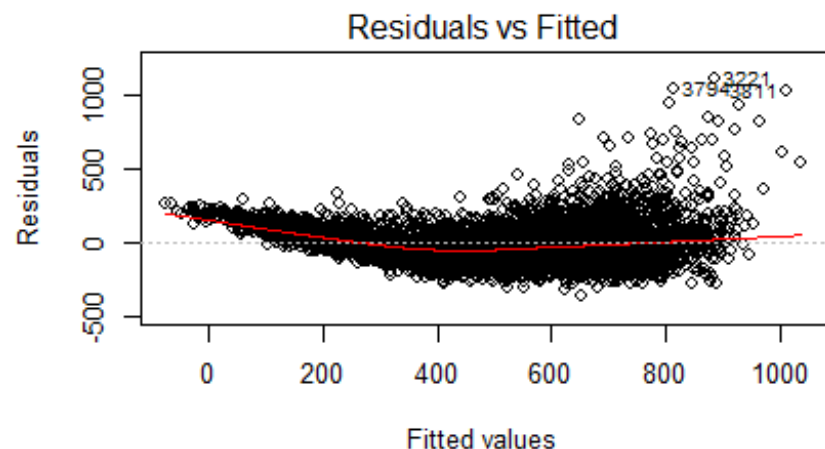
Multiple R-squared: 0.7283, Adjusted R-squared: 0.7282

F-statistic: 7454 on 2 and 5561 DF, p-value: < 2.2e-16

Interpretação: O modelo em conjunto é estatisticamente significativo pois o p-value é menor do 5%.

Modelos de Regressão

- Gráfico dos resíduos



Modelos de Regressão

- **Aplicação em R**

- Consideremos o modelo:

$$\text{Salário} = \beta_0 + \beta_1 \text{AnosEstudo} + \beta_2 \text{Experiencia} + \beta_3 \text{TempoServico} + \varepsilon$$

onde *Salário* = Salário horário, *anos de estudo* = anos de estudo em anos;
experiência = experiência em anos; *Tempo de serviço* = tempo de serviços em anos.

- *Importe os dados da planilha: salario.xls;*
- *Quais seriam os sinais esperados para β_1 , β_2 e β_3 ;*
- *Que variável explanatória teria mais influência no salário;*
- *Estime o modelo acima;*
- *Interprete os resultados dos coeficientes estimados, interprete o grau de ajuste do modelo aos dados. Teste a significância individual e coletiva dos coeficientes estimados;*

Modelos de Regressão

```
> mod2 <- lm(salario ~ experiencia + temposervico  
+            , data = dados2)  
> summary(mod2)
```

Call:

```
lm(formula = salario ~ experiencia + temposervico, data = dados2)
```

Residuals:

| Min | 1Q | Median | 3Q | Max |
|--------|--------|--------|-------|--------|
| -9.015 | -2.056 | -1.014 | 1.129 | 16.567 |

Coefficients:

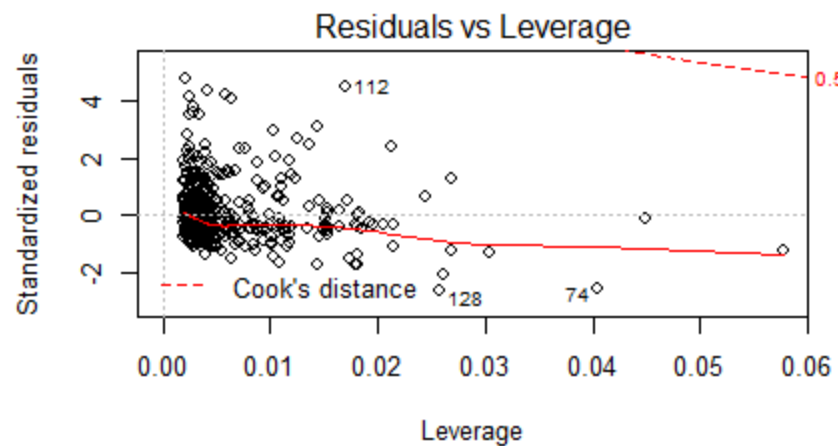
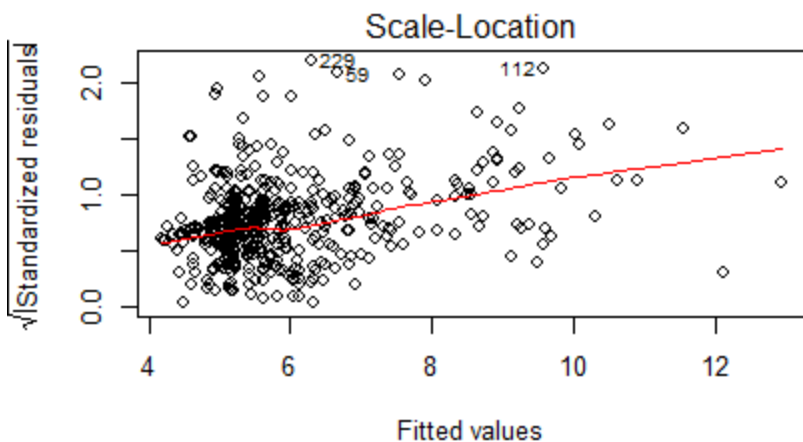
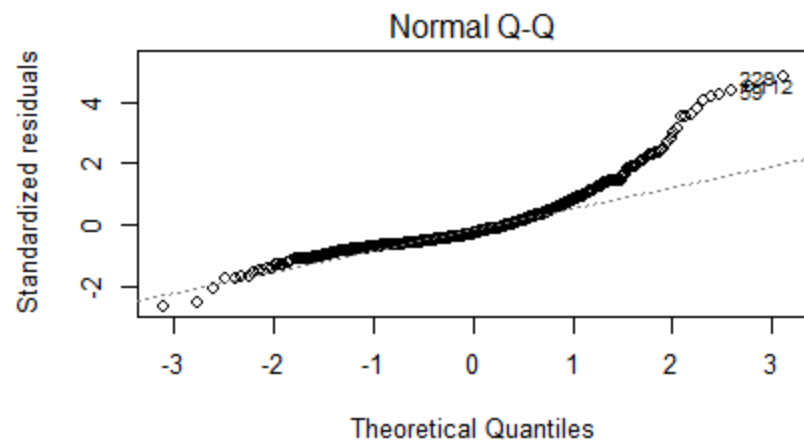
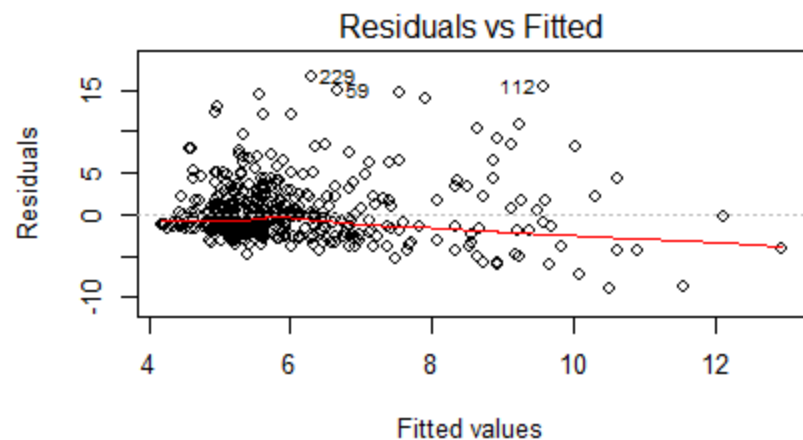
| | Estimate | Std. Error | t value | Pr(> t) |
|--------------|----------|------------|---------|----------|
| (Intercept) | 5.25819 | 0.24255 | 21.678 | < 2e-16 |
| experiencia | -0.02186 | 0.01284 | -1.702 | 0.0894 |
| temposervico | 0.19783 | 0.02413 | 8.198 | 1.89e-15 |

Residual standard error: 3.461 on 523 degrees of freedom

Multiple R-squared: 0.1252, Adjusted R-squared: 0.1218

F-statistic: 37.42 on 2 and 523 DF, p-value: 6.492e-16

Modelos de Regressão



Obrigado.