

# O ensino do R aplicado à Economia através do R Markdown

Segundo R Day - 17:10h.

Marcos Minoru Hasegawa

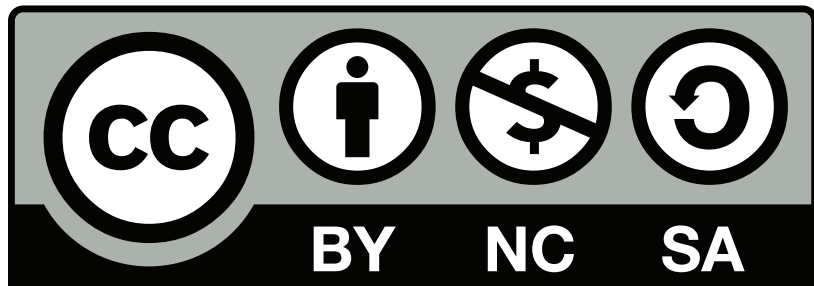
13/12/2019

# Licença deste material

## **Licença Creative Commons BY NC SA**

Este trabalho está licenciado sob a Licença Atribuição - Não Comercial - Compartilha Igual 4.0 Internacional Creative Commons. Para visualizar uma cópia desta licença, visite <http://creativecommons.org/licenses/by-nc-sa/4.0/> ou mande uma carta para Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

Licença deste material



## Segundo R Day

**Data:** 13 dezembro de 2019

**Local:** Auditório do Prédio da Administração do Centro Politécnico da UFPR

**Horário:** 9:00h. - 18:00h.

**Organização e realização:** Laboratório de Estatística e Geoinformação - LEG do Departamento de Estatística da UFPR.

## Breve Apresentação Pessoal

- ▶ Engenheiro Agrônomo - UNESP, Mestre em Economia Agrária - ESALQ/USP, Doutor em Economia Aplicada - ESALQ/USP
- ▶ Professor Adjunto do Departamento de Economia do Setor de Ciências Sociais Aplicadas da Universidade Federal do Paraná.
- ▶ Um dos professores responsáveis pelas disciplinas SE305 Estatística Econômica e Introdução à Economia e SE308 Econometria no curso de graduação em Economia.
- ▶ Áreas de interesses: Economia Agrícola, Ciência Regional e Métodos Quantitativos Aplicados à Economia.
- ▶ Usuário em nível intermediário: linguagem R,  $\text{\LaTeX}$ , SAS, Gempack.
- ▶ e-mail: [hasegawa@ufpr.br](mailto:hasegawa@ufpr.br)

## Antes de descobrir o **R Markdown**

- ▶ MS-Word, MS-Powerpoint, MS-Excel e programas estatísticos e econométricos proprietários.
- ▶  $\text{\LaTeX}$  e R separados.
- ▶ Juntando  $\text{\LaTeX}$  e R através do **Sweave**, mas restrito ao material em PDF.

# Usando o R Markdown

- ▶  $\text{\LaTeX}$  e R juntos através do **R Markdown**
- ▶ **arquivo .Rmd**: é um registro de sua pesquisa que contém o código que o cientista precisa para reproduzir o seu trabalho ao longo da narrativa que leitor necessita para compreender o teu trabalho.
- ▶ **pesquisa reproduzível**: a um click ou a digitação de um comando, é possível executar novamente o código em um arquivo **R Markdown** para reproduzir o seu trabalho e exportar os resultados em um relatório elaborado.
- ▶ **documentos dinâmicos**: dinâmico no sentido de permitir escolher uma variedade de formatos para o qual será exportado o relatório elaborado, formatos que incluem html, pdf, MS-Word, rtf, slides baseados em html ou pdf, etc.

# Um pouco de história sobre o que é hoje o **R Markdown**

- ▶ O formato de documento **R Markdown** inicialmente introduzido no pacote **knitr** no início de 2012 por **Yihui Xie**.
- ▶ A ideia era executar pedaços de códigos de R ou outras linguagens embutidos nos documentos **Markdown** (html).
- ▶ A versão original do **Markdown**, desenvolvido por **John Gruber** a partir de 2004 com a parceria de **Aaron Swartz**, tornou-se popular pela sua simplicidade.
- ▶ Mas o **Markdown** não era adequado para escrever documentos altamente técnicos, pois não existia sintaxe para tabelas, notas de rodapé, expressões matemáticas ou citações.
- ▶ A partir de 2006, John MacFarlane começou a desenvolver o pacote **Pandoc** para converter documentos **Markdown** para uma grande variedade de formatos.



## O R Markdown em resumo

Resumindo:

- ▶ O **R Markdown** tem como base os pacotes **knitr** e **Pandoc**.
- ▶ o **knitr** executa o código da linguagem de programação R (ou python, Julia, C++, etc.) embutida no **Markdown** e converte **R Markdown** para **Markdown**.
- ▶ o **Pandoc** renderiza o **Markdown** para um formato de saída que deseje tais como: pdf, html, MS-Word, MS-Powerpoint, etc.

# Pacote **rmarkdown**

Criado no início de 2014, permite realizar um grande número de tarefas com **R Markdown**:

- ▶ Compilar um documento R Markdown para gerar um relatório em diferentes formatos: pdf, html, MS-Word, etc.
- ▶ Fazer slides para apresentações: html5,  $\text{\LaTeX}$  Beamer ou MS-Powerpoint.
- ▶ Produzir *dashboards* com layouts atrativos, interativos e flexíveis.
- ▶ Construir aplicações interativas baseado no **Shiny**.
- ▶ Escrever artigos para revistas científicas.
- ▶ Escrever livros com múltiplos capítulos.
- ▶ Gerar websites e blogs.

# Livros sobre R Markdown

Livros no formato digital **GRÁTIS** na internet

- ▶ Xie, Y. Dynamic Documents with R and Knitr. 2nd ed. Boca Raton: Chapman Hall/CRC, 2015
- ▶ Xie, Y. Bookdown: Authoring Books and Technical Documents with R Markdown. Boca Raton: Chapman Hall/CRC, 2016.
- ▶ Xie, Y.; Allaire, J.J.; Golemund G. R Markdown: The Definitive Guide. Boca Raton: Chapman Hall/CRC. 2018

# O RStudio IDE

- ▶ É um ambiente de desenvolvimento integrado (IDE) para o R.
- ▶ Inclui console, editor com destacamento da sintaxe que suporta execução direta de código, bem como ferramentas para o gerenciamento de gráficos, históricos, depuração e área de trabalho.
- ▶ Versões código aberto (AGPL v.3) e comercial.
- ▶ versões desktop (Windows, Mac e Linux) e servidor (Linux Debian/Ubuntu, Red Hat/CentOS e SUSE).
- ▶ <https://rstudio.com/>

# O R Markdown na UFPR

- ▶ Formatação padrão em R Markdown para TCC, monografia, dissertação e tese nas normas da UFPR (ou será que existe e eu não sei??)
- ▶ Para o  $\text{\LaTeX}$  puro existe a iniciativa do Emílio E. Kawamura que desenvolveu um template que serve de modelo para TCC, Monografia, Dissertação e Tese adaptado para as normas ABNT e UFPR.
  - ▶ <https://pt.overleaf.com/latex/templates/abntex-ufpr/bsswfjqykhhd>
  - ▶ <https://github.com/EEKBR/ufpr-abntex>
- ▶ Usuários do R Markdown dispersos e isolados na UFPR
- ▶ Resistência dos colegas professores de sair da sua zona de conforto, mesmo só para o uso do R.
- ▶ **R Markdown** >  $\text{\LaTeX}$  + R

# O RStudio Cloud

Se tem problemas:

- ▶ para instalar o R, o RStudio, o  $\text{\LaTeX}$ , o python, etc.
- ▶ de equipamentos no laboratório com configurações não muito interessantes.

A solução pode ser:

## ▶ **RStudio Servidor**

- ▶ grátis se não precisar de suporte técnico e algumas funcionalidades.
- ▶ necessidade de equipamentos não muito baratos, além da infraestrutura.

## ▶ **RStudio Cloud:**

- ▶ em fase de experimentação e melhoria.
- ▶ Grátis por enquanto.

# O RStudio Cloud

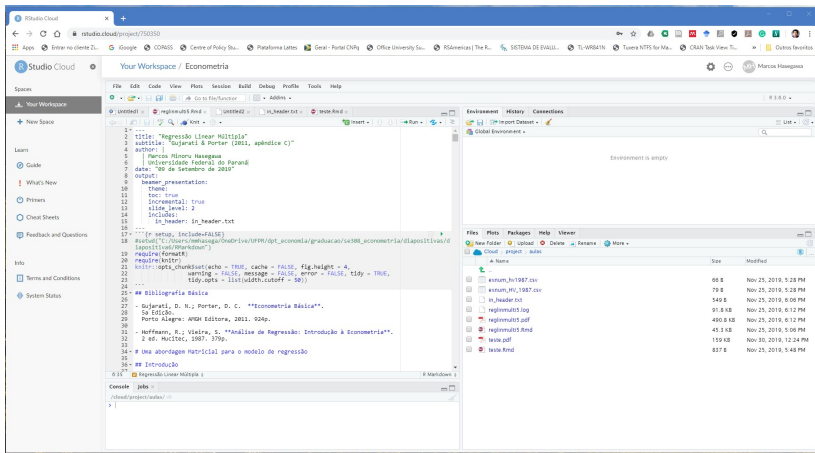


Figura 1: O meu RStudio Cloud no navegador Google Chrome.

# Ensinando o R aplicado a Economia através do **R Markdown**

Para finalizar, é apresentada uma sequência de slides mostrando como o **R Markdown** ajuda a dar uma outra dinâmica nas minhas aulas de Econometria no curso de Economia.



## Estimativa por M.Q.O.

Para obter uma estimativa M.Q.O. de  $\beta$ , primeiro escreve-se a função de regressão amostral de  $k$  variáveis:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \dots + \hat{\beta}_k X_{ki} + \hat{u}_i \quad (1)$$

que pode ser escrita de forma mais compacta

$$\mathbf{y} = \mathbf{X}\hat{\beta} + \hat{\mathbf{u}} \quad (2)$$

## Estimativa por M.Q.O.

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{21} & X_{31} & \dots & X_{k1} \\ 1 & X_{22} & X_{32} & \dots & X_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{2n} & X_{3n} & \dots & X_{kn} \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} + \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_n \end{bmatrix} \quad (3)$$

$$\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}}$$

onde

- ▶  $\hat{\boldsymbol{\beta}}$  é um vetor coluna de  $k$  elementos dos estimadores de M.Q.O. dos coeficientes de regressão e;
- ▶  $\hat{\mathbf{u}}$  é um vetor coluna  $n \times 1$  de  $n$  resíduos.

## Estimativa por M.Q.O.

No caso de  $k$  variáveis os estimadores de M.Q.O. são obtidos ao minimizar

$$\text{Min} \sum \hat{u}_i^2 = \sum \left( Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \dots - \hat{\beta}_{ki} X_{ki} \right)^2 \quad (4)$$

onde  $\sum \hat{u}_i^2$  é a soma dos quadrados dos resíduos (SQR). Em notação matricial equivale a minimizar  $\hat{\mathbf{u}}^T \hat{\mathbf{u}}$  pois

$$\hat{\mathbf{u}}^T \hat{\mathbf{u}} = \begin{bmatrix} \hat{u}_1 & \hat{u}_2 & \dots & \hat{u}_n \end{bmatrix} \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_n \end{bmatrix} = \hat{u}_1^2 + \hat{u}_2^2 + \dots + \hat{u}_n^2 = \sum \hat{u}_i^2 \quad (5)$$

## Estimativa por M.Q.O.

A partir de (2), obtemos

$$\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}. \quad (6)$$

Portanto,

$$\begin{aligned} \hat{\mathbf{u}}^T \hat{\mathbf{u}} &= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ \hat{\mathbf{u}}^T \hat{\mathbf{u}} &= (\mathbf{y}^T \mathbf{y} - 2\hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{y} + \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}}). \end{aligned} \quad (7)$$

## Estimativa por M.Q.O.

Calculando a derivada parcial de  $\hat{\mathbf{u}}^T \hat{\mathbf{u}}$  temos

$$\frac{\partial \hat{\mathbf{u}}^T \hat{\mathbf{u}}}{\partial \hat{\boldsymbol{\beta}}} = -2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}}$$

igualando a zero pelas condições de primeira ordem

$$(\mathbf{X}^T \mathbf{X}) \hat{\boldsymbol{\beta}} = \mathbf{X}^T \mathbf{y} \quad (8)$$

pré-multiplicando por  $(\mathbf{X}^T \mathbf{X})^{-1}$

$$(\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{X}) \hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

## Estimativa por M.Q.O.

Lembre-se que  $(\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{X}) = \mathbf{I}$  e por isso

$$\begin{aligned}\mathbf{I} \hat{\boldsymbol{\beta}} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \\ \hat{\boldsymbol{\beta}}_{k \times 1} &= (\mathbf{X}^T \mathbf{X})_{k \times k}^{-1} \mathbf{X}^T_{k \times n} \mathbf{y}_{n \times 1}\end{aligned}\tag{9}$$

## Exemplo numérico extraído de Hoffmann e Vieira (1987)

Tomando o exemplo numérico apresentado em Hoffmann e Vieira (1987), considerando o seguinte modelo econométrico:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

sendo:

- ▶  $Y_i$  é a variável dependente;
- ▶  $X_{2i}$  e  $X_{3i}$  são as variáveis explicativas e;
- ▶  $u_i$  o termo estocástico.

## Exemplo numérico extraído de Hoffmann e Vieira (1987)

Matricialmente o modelo econométrico fica

$$\underset{6 \times 1}{\mathbf{y}} = \underset{6 \times 3}{\mathbf{X}} \underset{3 \times 1}{\beta} + \underset{6 \times 1}{\mathbf{u}}$$



## Exemplo numérico extraído de Hoffmann e Vieira (1987)

Colocando o exemplo numérico na forma matricial temos

$$\begin{bmatrix} 1,5 \\ 6,5 \\ 10 \\ 11 \\ 11,5 \\ 16,5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 2 \\ 1 & 1 & 4 \\ 1 & 2 & 2 \\ 1 & 2 & 4 \\ 1 & 3 & 6 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \end{bmatrix}$$

## Usando o R: montando o vetor y

```
y <- matrix(c(1.5, 6.5, 10, 11, 11.5, 16.5), ncol = 1,  
            nrow = 6, byrow = T)
```

y

```
##      [,1]
```

```
## [1,]  1.5
```

```
## [2,]  6.5
```

```
## [3,] 10.0
```

```
## [4,] 11.0
```

```
## [5,] 11.5
```

```
## [6,] 16.5
```

## Usando o R: montando a matriz X

```
X <- matrix(c(1, 1, 1, 1, 1, 1, 0, 1, 1, 2, 2, 3, 0,  
             2, 4, 2, 4, 6), ncol = 3, nrow = 6, byrow = F)  
X
```

```
##      [,1] [,2] [,3]  
## [1,]    1    0    0  
## [2,]    1    1    2  
## [3,]    1    1    4  
## [4,]    1    2    2  
## [5,]    1    2    4  
## [6,]    1    3    6
```

## Exemplo numérico extraído de Hoffmann e Vieira (1987)

Assim temos

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix}$$

## Exemplo numérico extraído de Hoffmann e Vieira (1987)

$$\mathbf{X}^T \mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ X_{21} & X_{22} & X_{23} & \dots & X_{2\ 6} \\ X_{31} & X_{32} & X_{33} & \dots & X_{3\ 6} \end{bmatrix} \begin{bmatrix} 1 & X_{21} & X_{31} \\ 1 & X_{22} & X_{32} \\ 1 & X_{23} & X_{33} \\ \vdots & \vdots & \vdots \\ 1 & X_{2\ 6} & X_{3\ 6} \end{bmatrix} =$$

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 2 & 2 & 3 \\ 0 & 2 & 4 & 2 & 4 & 6 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 2 \\ 1 & 1 & 4 \\ 1 & 2 & 2 \\ 1 & 2 & 4 \\ 1 & 3 & 6 \end{bmatrix}$$

## Usando R: para calcular $X^T X$

```
Xt <- t(X)
```

```
Xt
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6]
## [1,]    1    1    1    1    1    1
## [2,]    0    1    1    2    2    3
## [3,]    0    2    4    2    4    6
```

```
XtX <- Xt %*% X
```

```
XtX
```

```
##      [,1] [,2] [,3]
## [1,]    6    9   18
## [2,]    9   19   36
## [3,]   18   36   76
```

## Exemplo numérico extraído de Hoffmann e Vieira (1987)

e

$$\mathbf{X}^T \mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ X_{21} & X_{22} & X_{23} & \dots & X_{26} \\ X_{31} & X_{32} & X_{33} & \dots & X_{36} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_6 \end{bmatrix} =$$
$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 2 & 2 & 3 \\ 0 & 2 & 4 & 2 & 4 & 6 \end{bmatrix} \begin{bmatrix} 1,5 \\ 6,5 \\ 10 \\ 11 \\ 11,5 \\ 16,5 \end{bmatrix}$$

## Usando R: para calcular $\mathbf{X}^T\mathbf{y}$

```
Xty <- Xt %*% y  
Xty
```

```
##      [,1]  
## [1,]   57  
## [2,]  111  
## [3,]  220
```



## O cálculo da inversa de $\mathbf{X}^\top \mathbf{X}$

$$(\mathbf{X}^\top \mathbf{X})^{-1} = \frac{1}{|\mathbf{X}^\top \mathbf{X}|} \mathbf{Cof} (\mathbf{X}^\top \mathbf{X})^\top$$

onde

$$\mathbf{Cof} = (-1)^{i+j} |\mathbf{M}_{ij}|$$

## Usando R: para calcular a inversa de $\mathbf{X}^T\mathbf{X}$

```
XtXinv <- solve(XtX)
```

```
XtXinv
```

```
##           [,1]    [,2]    [,3]
## [1,]  0.6166667 -0.150 -0.0750
## [2,] -0.1500000  0.550 -0.2250
## [3,] -0.0750000 -0.225  0.1375
```

## Usando R: cálculo do vetor $\hat{\beta}$

```
beta <- XtXinv %*% Xty  
beta
```

```
##      [,1]  
## [1,]    2  
## [2,]    3  
## [3,]    1
```

## Estimação da regressão usando a função `lm()`

Depois de gerar o CSV separado por vírgulas dos dados do exemplo do Hoffmann e Vieira (1987) utiliza-se a função `read.csv2()` para leitura e salvamento dos dados em formato R.

```
HV1987 <- read.csv2(file = "exnum_HV_1987.csv",  
                    header = TRUE)  
head(HV1987)
```

##		y	X2	X3
##	1	1.5	0	0
##	2	6.5	1	2
##	3	10.0	1	4
##	4	11.0	2	2
##	5	11.5	2	4
##	6	16.5	3	6

## Estimação da regressão usando a função `lm()`

- ▶ Para realizar os cálculos da estimação da regressão linear, usa-se a função `lm()` cujo pacote já foi instalado na ocasião da instalação do R.
- ▶ O objeto ou variável que contém os dados devem estar no formato `data.frame` e não no formato matricial.

# Estimação da regressão usando a função `lm()`

Usando a função `lm()`:

```
hv <- lm(y ~ X2 + X3, data = HV1987)
```

```
hv
```

```
##
```

```
## Call:
```

```
## lm(formula = y ~ X2 + X3, data = HV1987)
```

```
##
```

```
## Coefficients:
```

## (Intercept)	X2	X3
## 2	3	1

## Estimação da regressão usando a função **lm()**

obtendo o resumo da estimação com a função **summary()**

```
hvs <- summary(hv)
```

## Estimação da regressão usando a função `lm()`

```
##  
## Call:  
## lm(formula = y ~ X2 + X3, data = HV1987)  
##  
## Residuals:  
##      1      2      3      4      5      6  
## -0.5 -0.5  1.0  1.0 -0.5 -0.5  
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)  
## (Intercept)   2.0000     0.7853   2.547   0.0842 .  
## X2             3.0000     0.7416   4.045   0.0272 *  
## X3             1.0000     0.3708   2.697   0.0740 .  
## ---  
## Signif. codes:  
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 1 on 3 degrees of freedom
```



## Estimação da regressão usando a função `lm()`

```
Call:
lm(formula = y ~ x2 + x3, data = HV1987)

Residuals:
    1     2     3     4     5     6 
-0.5 -0.5  1.0  1.0 -0.5 -0.5 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.0000     0.7853   2.547   0.0842 .
x2             3.0000     0.7416   4.045   0.0272 *
x3             1.0000     0.3708   2.697   0.0740 .
---
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1 on 3 degrees of freedom
(4 observations deleted due to missingness)
Multiple R-squared:  0.9767,    Adjusted R-squared:  0.9611 
F-statistic: 62.75 on 2 and 3 DF,  p-value: 0.003567
```

Figura 2: Resultado da função `summary()`

## Bibliografia Consultada

- ▶ Gujarati, D. N.; Porter, D. C. **Econometria Básica**. 5a Edição. Porto Alegre: AMGH Editora, 2011. 924p.
- ▶ Hoffmann, R.; Vieira, S. **Análise de Regressão: Introdução à Econometria**. 2 ed. Hucitec, 1987. 379p.
- ▶ Xie, Y. **Dynamic Documents with R and Knitr**. 2nd ed. Boca Raton: Chapman Hall/CRC, 2015
- ▶ Xie, Y. **Bookdown: Authoring Books and Technical Documents with R Markdown**. Boca Raton: Chapman Hall/CRC, 2016.
- ▶ Xie, Y.; Allaire, J.J.; Golemund G. **R Markdown: The Definitive Guide**. Boca Raton: Chapman Hall/CRC, 2018