

# Predições para Gestão de Projetos com Regressão Linear Múltipla em R

Equipe EducaECO

13/08/2025

```
knitr::opts_chunk$set(echo = FALSE)
library("dplyr")
# Bibliotecas necessárias
library(magrittr)
library(stringr)
library(tidyverse)

#diretório do projeto
setwd("c:/R/gestao")

#ler as sprints
dfsprint=read.csv("sprints.csv",sep=";",fileEncoding="latin1")
#ler as atividades
dfativ=read.csv("atividades.csv",sep=";",fileEncoding="latin1")

#separa os dados de treino da trilha 1
dfsprint_treino<-dfsprint[1:5,]
#separa os dados para a predição da trilha 1
dfsprint_teste<-dfsprint[6:12,]

#separa os dados de treino trilha 2
dfativ_treino<-dfativ[1:17,]
#separa os dados para a predição trilha 2
dfativ_teste<-dfativ[18:18,]

# Ajustando o modelo de regressão linear múltipla trilha 1
modelo <- lm(horas_reais ~ story_point + complexidade + horas_estimadas, data = dfsprint_treino)

# Ajustando o modelo de regressão linear múltipla trilha 2
modelo2 <- lm(horas_reais ~ story_point + complexidade + dependencia + interrupcao + horas_estimadas, data = dfativ_treino)

#verificando a qualidade do modelo trilha 1
```

```
summary(modelo)$r.squared
```

```
## [1] 0.9956681
```

```
summary(modelo)$adj.r.squared
```

```
## [1] 0.9826723
```

```
#verificando a qualidade do modelo trilha 2  
summary(modelo2)$r.squared
```

```
## [1] 0.9833852
```

```
summary(modelo2)$adj.r.squared
```

```
## [1] 0.975833
```

```
#faz a predição da trilha 1  
dfsprint_teste$horas_previstas <- predict(modelo, newdata = dfsprint_teste)
```

```
#faz a predição da trilha 2  
dfativ_teste$horas_previstas <- predict(modelo, newdata = dfativ_teste)
```

```
#ajustando data_frame para exibir os desvios trilha 1  
dfsprint_treino$horas_previstas=NA  
dfsprint_treino$desvio=(dfsprint_treino$horas_reais - dfsprint_treino$horas_estimadas) / dfsprint_treino$horas_previstas
```

```
dfsprint_teste$desvio=(dfsprint_teste$horas_previstas - dfsprint_teste$horas_estimadas) / dfsprint_teste$horas_previstas
```

```
#desvios trilha 2  
desvio_ativ<-(dfativ_teste$horas_previstas - dfativ_teste$horas_estimadas) / dfativ_teste$horas_previstas
```

```
#data frame para exibição dos desvios  
df_desvio<-rbind(dfsprint_treino, dfsprint_teste)
```

```
#ajuste para gerar gráfico  
dfgrafico<- df_desvio %>% pivot_longer(  
  cols = c(horas_reais, horas_previstas, horas_estimadas), # as colunas desse intervalo  
  names_to = "tipo_horas", # terão seus nomes armazenados nessa nova coluna  
  values_to = "qtd_horas") # e os seus valores armazenados nessa nova coluna
```

## Predição do tempo das Sprints

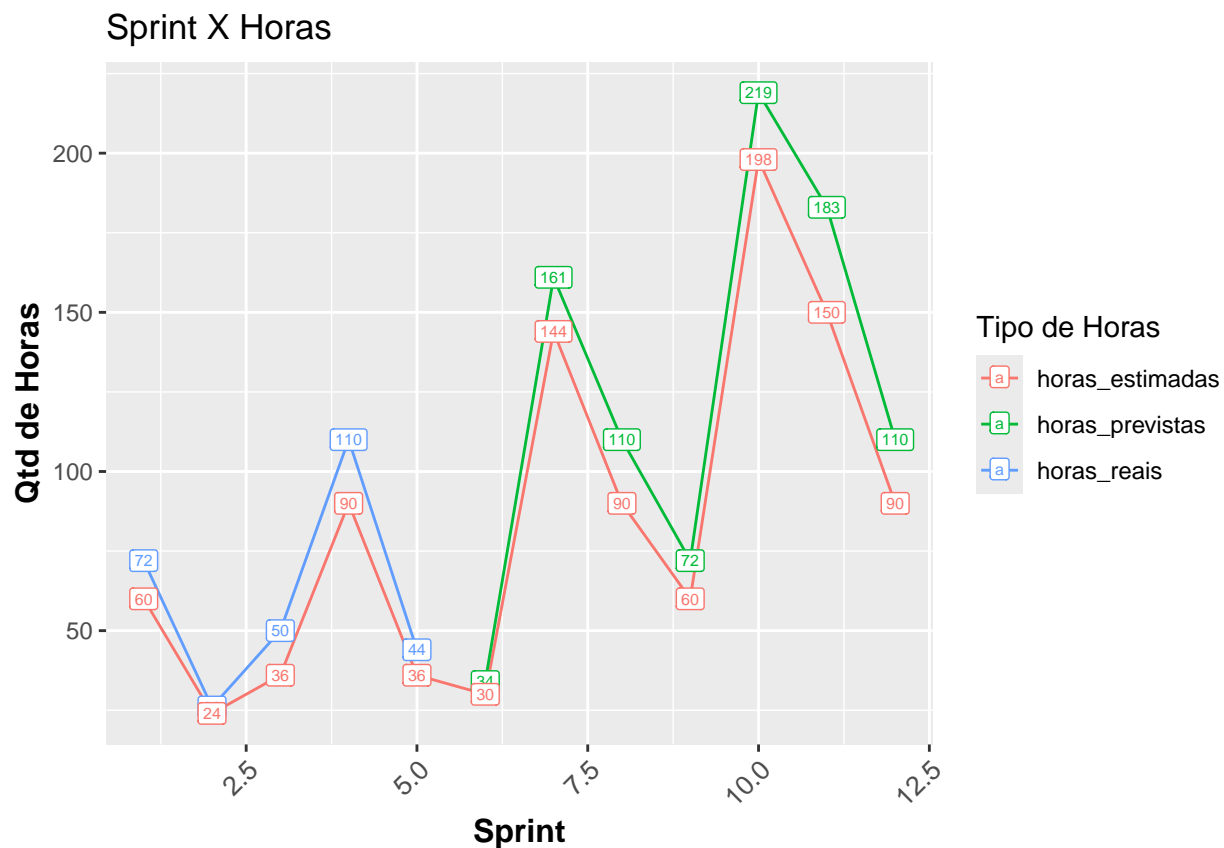
Projeto com 12 Sprints — 5 concluídas (com desvios) + 5 futuras (com predição)

Sprints 1–5 concluídas com problemas entre horas estimadas e reais. Treinamos um modelo com a função `lm()` do R base nas sprints 1–5. Prevemos as horas reais para as sprints 6–12 e comparamos com o estimado. Métricas:  $R^2$  e  $R^2$  Ajustado. Gráfico: Estimado x Real/Previsto por sprint.

Table 1: Desvios Real/Previsto X Estimado (horas) por Sprint

sprint	story_point	horas_estimadas	horas_reais	horas_previstas	desvio
1	10	60	72	NA	20.000000
2	8	24	26	NA	8.333333
3	6	36	50	NA	38.888889
4	10	90	110	NA	22.222222
5	6	36	44	NA	22.222222
6	10	30	NA	34	13.333333
7	16	144	NA	161	11.805556
8	10	90	NA	110	22.222222
9	10	60	NA	72	20.000000
10	26	198	NA	219	10.606061
11	26	150	NA	183	22.000000
12	10	90	NA	110	22.222222

Gráfico — Estimado x Real/Previsto (horas) por Sprint



## Forma de verificar a qualidade do modelo

### R-quadrado

O R-quadrado é uma medida estatística de quão próximos os dados estão da linha de regressão ajustada.

Ele também é conhecido como o coeficiente de determinação ou o coeficiente de determinação múltipla para a regressão múltipla.

A definição do R-quadrado é bastante simples: é a porcentagem da variação da variável resposta que é explicada por um modelo linear. Ou:

$$\text{R-quadrado} = \text{Variação explicada} / \text{Variação total}$$

O R-quadrado está sempre entre 0 e 100%:

0% indica que o modelo não explica nada da variabilidade dos dados de resposta ao redor de sua média. 100% indica que o modelo explica toda a variabilidade dos dados de resposta ao redor de sua média.

R-quadrado ajustado

O R-quadrado ajustado (ou  $R^2$  ajustado) é uma métrica estatística que avalia a qualidade de um modelo de regressão, levando em consideração o número de variáveis independentes (preditores) e o tamanho da amostra. Ele penaliza a inclusão de variáveis irrelevantes no modelo, ao contrário do R-quadrado simples, que sempre aumenta com a adição de novas variáveis.

## Predição do Tempo da Atividade 1 da Sprint 10 - Trilha 2

Sprints 1–5 concluídas com problemas entre horas estimadas e reais das atividades. Treinamos um modelo com a função `lm()` do R base das atividades das sprints 1–5. Prevemos as horas reais para a atividade 1 da sprint 10, que está prevista como a mais complexa da sprint, e comparamos com o estimado. Métricas:  $R^2$  e  $R^2$  Ajustado.

Table 2: Previsão da Atividade 1 da Sprint 10

	story_point	complexidade	dependencia	interrupcao	equipe	horas_estimadas	horas_previstas
18	8	3	1	0	2	55	80.25

**Desvio da Atividade:** 45.91 %