

checkpoint1

Rodolfo Viana

13-05-2015

Utilizando uma amostra dos gols do campeonato brasileiro de 2013, gostaríamos de saber se existe diferença significativa entre o momento em que o gol acontece no primeiro e segundo tempo das partidas. Utilizando a média do momento do gol obtemos os seguintes resultados:

```
library(dplyr)
library(ggplot2)

file <- read.csv("~/Projetos/DataAnalysis/Assignment3/brasileiro2013.csv", sep=";", header=TRUE)

fileSemNone <- file %>%
  filter(tempo_gol != "None")

fileSemNone$tempo_gol <- as.numeric(fileSemNone$tempo_gol)

primeiroTempo <- fileSemNone %>%
  filter(tempo_gol <= 45) %>%
  mutate(tempo_id = '1')

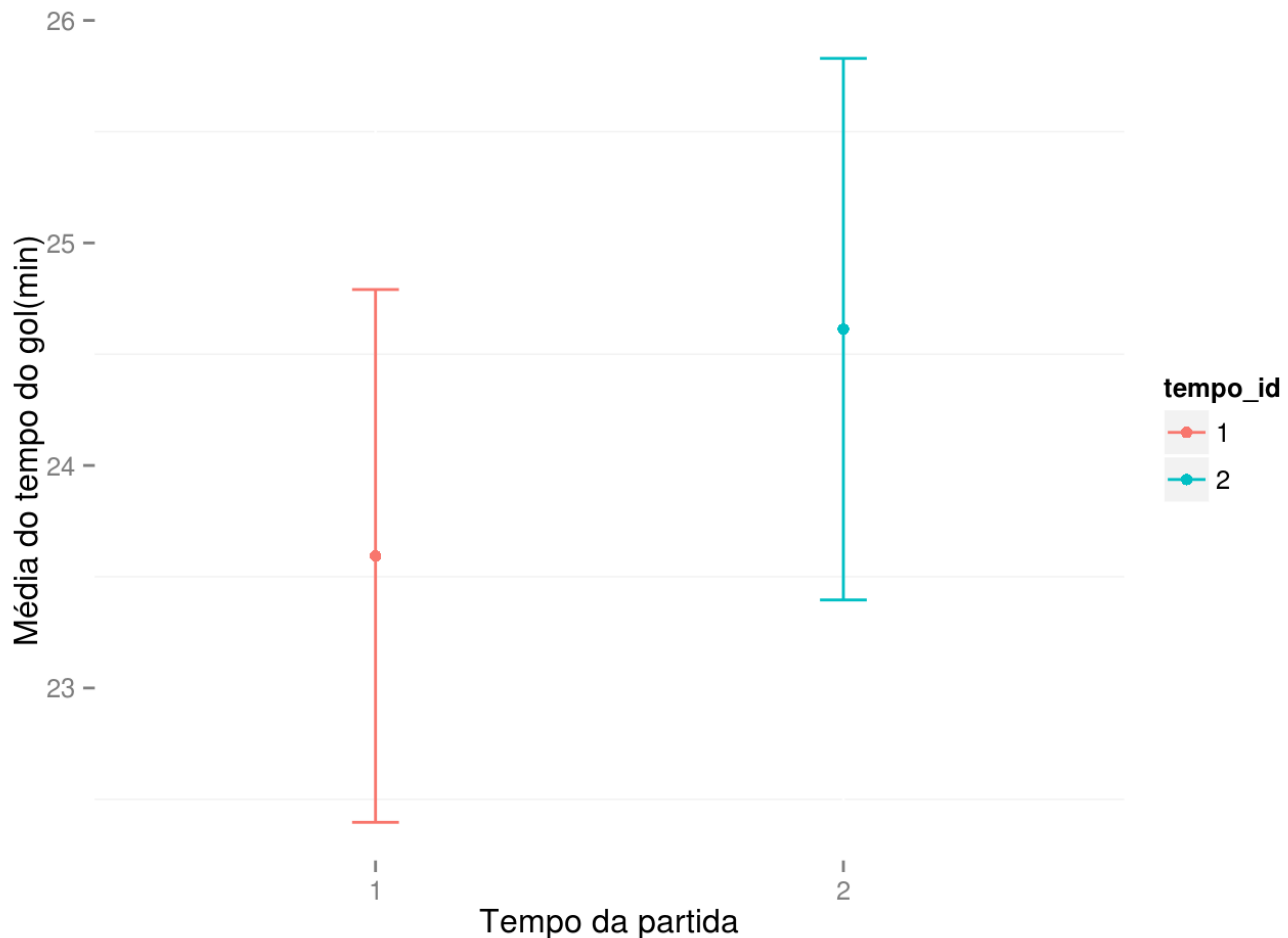
segundoTempo <- fileSemNone %>%
  filter(tempo_gol > 45) %>%
  mutate(tempo_id = '2')

segundoTempo$tempo_gol <- segundoTempo$tempo_gol - 45

se = sd(primeiroTempo$tempo_gol) / sqrt(nrow(primeiroTempo))
lower = mean(primeiroTempo$tempo_gol) - 1.96 * se
upper = mean(primeiroTempo$tempo_gol) + 1.96 * se
ic_1 = c(lower,upper)
se1 = sd(segundoTempo$tempo_gol) / sqrt(nrow(segundoTempo))
lower = mean(segundoTempo$tempo_gol) - 1.96 * se1
upper = mean(segundoTempo$tempo_gol) + 1.96 * se1
ic_2 = c(lower,upper)

samples = rbind(primeiroTempo,segundoTempo)

toPlot = summarise(group_by(samples, tempo_id), mean = mean(tempo_gol))
toPlot = mutate(toPlot, lower = ifelse(toPlot$tempo_id == 1,ic_1[1],ic_2[1]))
toPlot = mutate(toPlot, upper = ifelse(toPlot$tempo_id == 1,ic_1[2],ic_2[2]))
ggplot(toPlot, aes(x = tempo_id, y=mean, colour = tempo_id )) +
  geom_point() +
  geom_errorbar(aes(ymin=lower, ymax=upper), width=.1) +
  labs(y='Média do tempo do gol(min)', x='Tempo da partida') +
  theme(panel.background=element_blank())
```



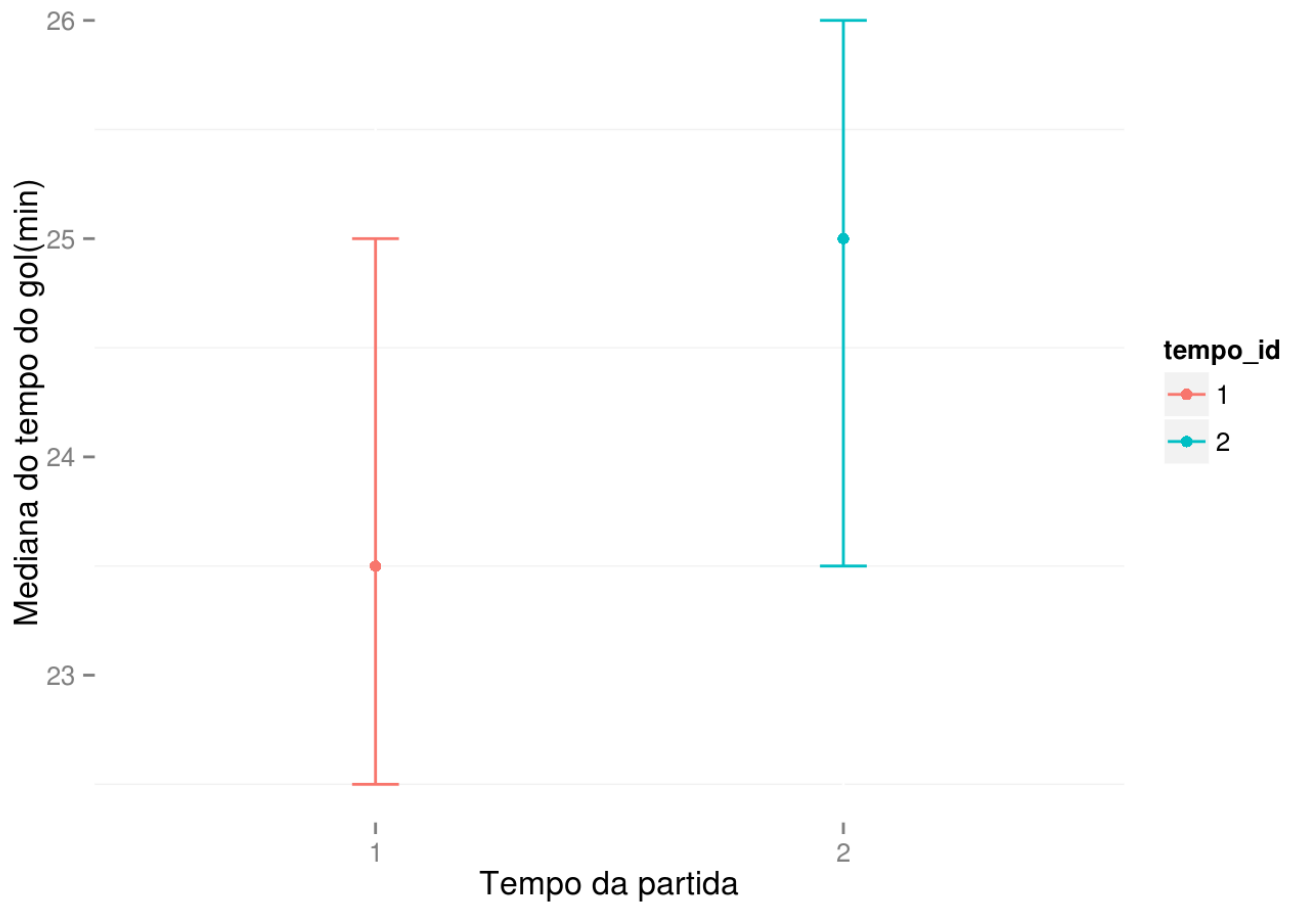
Olhando para o gráfico, podemos afirmar com 95% de confiança que a diferença entre o momento em que o gol acontece no primeiro e segundo tempo das partidas não é significativa.

Utilizando a mediana do momento do gol obtemos os seguintes resultados:

```
medianaPrimeiroTempo <- wilcox.test(primeiroTempo$tempo_gol, alternative = "two.sided", conf.int = TRUE)
medianaSegundoTempo <- wilcox.test(segundoTempo$tempo_gol, alternative = "two.sided", conf.int = TRUE)
jogos <- wilcox.test(primeiroTempo$tempo_gol, segundoTempo$tempo_gol, alternative = "two.sided", conf.int = TRUE)

ic_1 = c(medianaPrimeiroTempo$conf.int[1], medianaPrimeiroTempo$conf.int[2])
ic_2 = c(medianaSegundoTempo$conf.int[1], medianaSegundoTempo$conf.int[2])

toPlot = summarise(group_by(samples, tempo_id), median = median(tempo_gol))
toPlot = mutate(toPlot, lower = ifelse(toPlot$tempo_id == 1, ic_1[1], ic_2[1]))
toPlot = mutate(toPlot, upper = ifelse(toPlot$tempo_id == 1, ic_1[2], ic_2[2]))
ggplot(toPlot, aes(x = tempo_id, y=median, colour = tempo_id )) +
  geom_point() +
  geom_errorbar(aes(ymin=lower, ymax=upper), width=.1) +
  labs(y='Mediana do tempo do gol(min)', x='Tempo da partida') +
  theme(panel.background=element_blank())
```



Olhando para o gráfico, podemos afirmar com 95% de confiança que a diferença entre o momento em que o gol acontece no primeiro e segundo tempo das partidas não é significativa.

Como conclusão é possível notar que temos praticamente o mesmo resultado, tanto utilizando a média ou a mediana. Não existindo diferença na conclusão. Com os intervalos de confiança foi possível verificar mais claramente que para esse caso não existem diferenças entre o primeiro tempo e o segundo tempo em relação ao momento do gol.