

2) A partir das duas (2) variáveis adotadas para análise:

a) Desenvolva uma breve análise exploratória/descritiva das mesmas

```
library(dplyr)
library(knitr)

dados <- read.csv("brasileirao_serie_a.csv")

dados_filtrados <- dados %>%
  select(time = time_man, chutes = chutes_man, gols = gols_man) %>%
  filter(!is.na(chutes))

kable(head(dados_filtrados))
```

time	chutes	gols
Santos FC	3	1
Ceará SC	5	0
Atlético-PR	4	1
Ceará SC	4	0
EC Bahia	6	1
Atlético-MG	4	3

b) Desenvolva e **interprete de forma prática** uma análise de correlação.

c) Desenvolva e **interprete de forma prática** uma análise de regressão linear simples, incluindo a análise de resíduos e previsões para alguns valores estabelecidos para a variável independente, $X = x$.

```
library(ggplot2)
library(car)
library(ggpubr)

# Modelo de Regressao Linear Simples (MRLS)
mod <- lm(chutes ~ gols, data = dados_filtrados)

# Coeficientes Estimados
mod

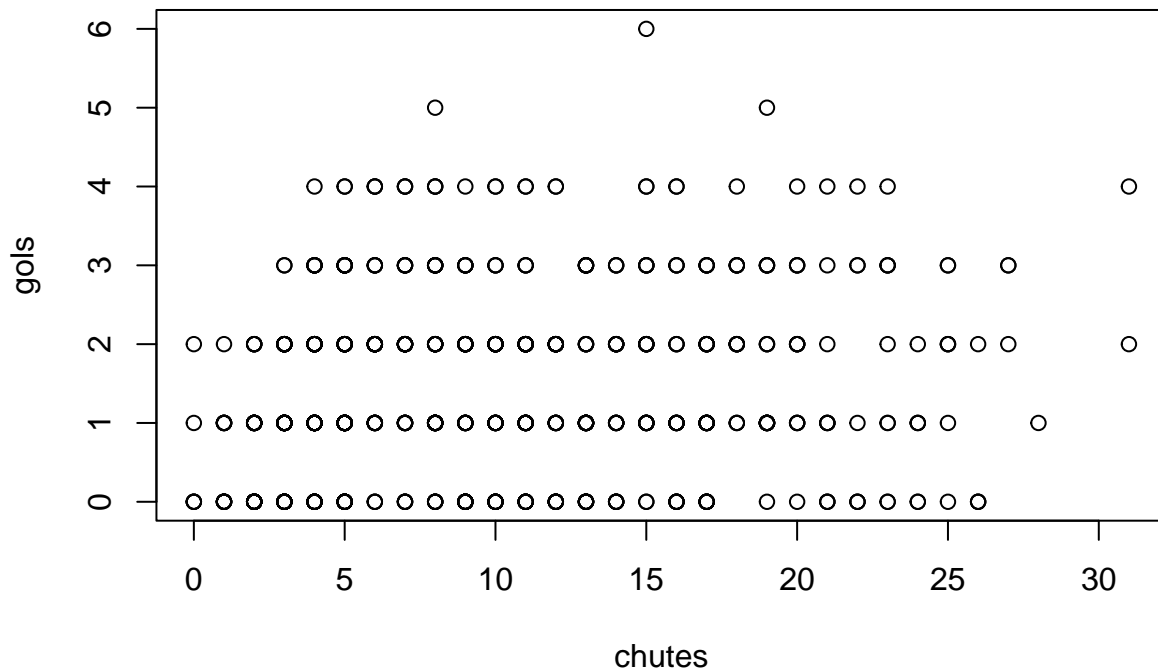
##
## Call:
## lm(formula = chutes ~ gols, data = dados_filtrados)
##
## Coefficients:
## (Intercept)      gols
##      8.511      0.977

# Inferências
summary(mod)
```

```
##
## Call:
## lm(formula = chutes ~ gols, data = dados_filtrados)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.465  -5.465  -1.442   4.489  20.535
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   8.5111     0.3598  23.655 < 2e-16 ***
## gols          0.9770     0.2038   4.794 1.97e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.178 on 751 degrees of freedom
## Multiple R-squared:  0.02969,    Adjusted R-squared:  0.0284
## F-statistic: 22.98 on 1 and 751 DF,  p-value: 1.974e-06
```

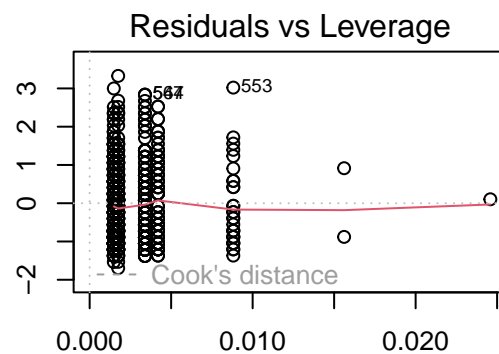
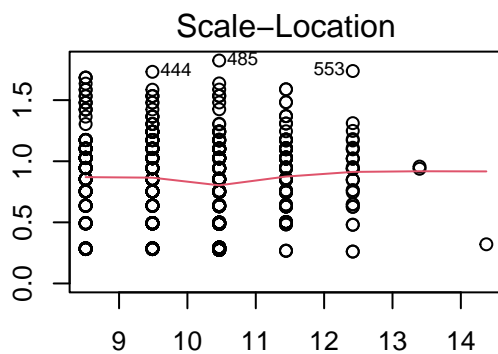
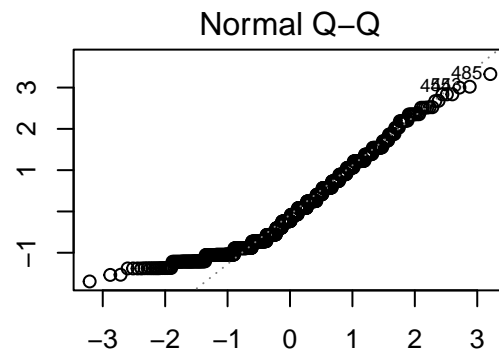
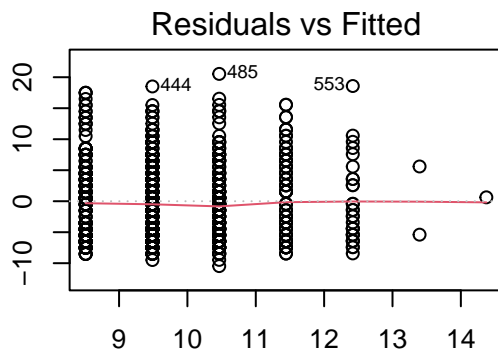
```
#Gráfico de relação
```

```
plot(dados_filtrados$chutes, dados_filtrados$gols, xlab="chutes", ylab="gols")
```



```
## Analise grafica baseada no modelo estimado
```

```
par(mfrow=c(2,2), mar=c(3,3,3,3))
plot(mod)
```



```
# Normalidade dos residuos:
shapiro.test(mod$residuals)
```

```
##
## Shapiro-Wilk normality test
##
## data: mod$residuals
## W = 0.93874, p-value < 2.2e-16
```

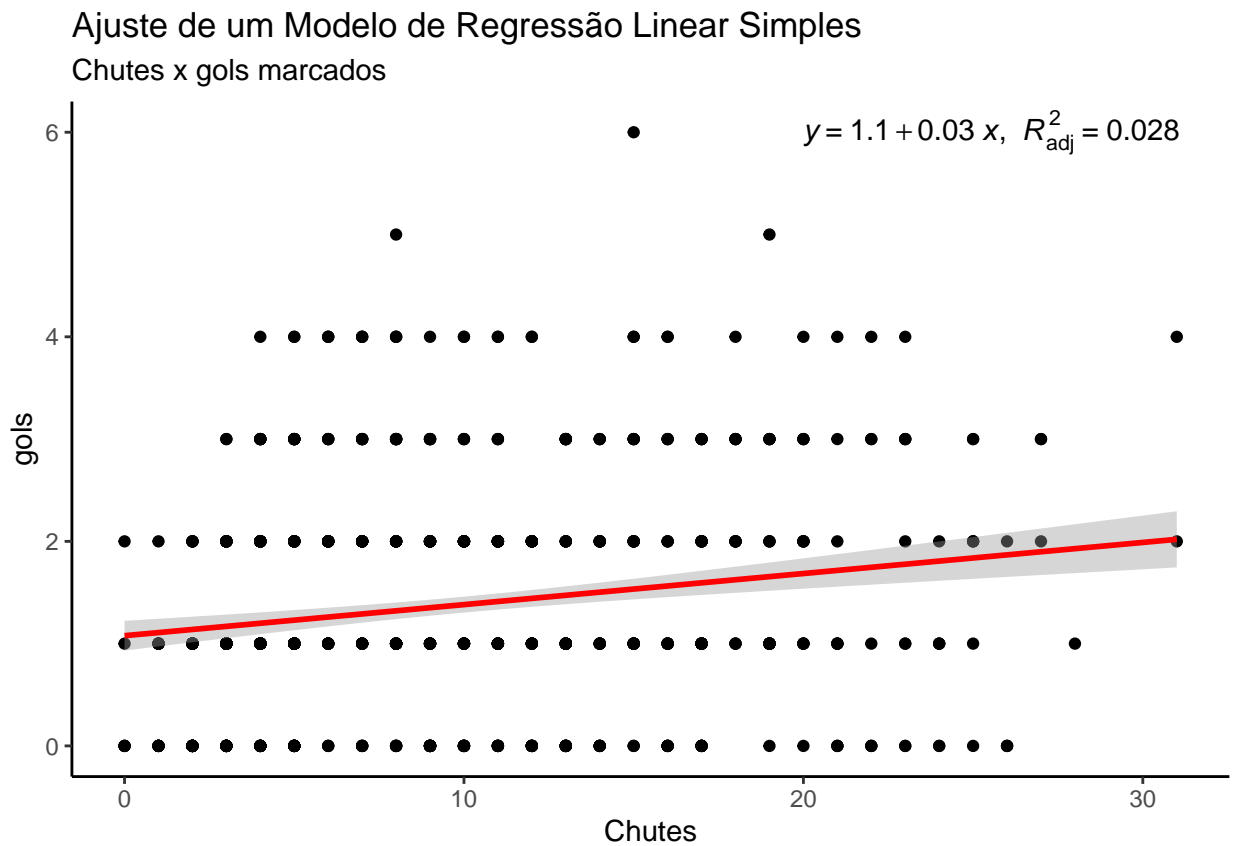
```
# Outliers nos residuos:
summary(rstandard(mod))
```

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
## -1.695425 -0.885386 -0.233911  0.000022  0.727833  3.326820
```

```
# Independencia dos residuos
durbinWatsonTest(mod)
```

```
## lag Autocorrelation D-W Statistic p-value
## 1 0.5278336 0.9413638 0
## Alternative hypothesis: rho != 0
```

```
dados_filtrados %>% ggplot(aes(x = chutes, y = gols)) +
  theme_classic() +
  geom_point() +
  geom_smooth(method = "lm", col = "red") +
  stat_regline_equation(aes(label = paste(..eq.label.., ..adj.rr.label..,
                                          sep = "*plain(\",\\\"~~\"))),
                      label.x = 20, label.y = 6) +
  labs(x='Chutes', y='gols',
       title='Ajuste de um Modelo de Regressão Linear Simples',
       subtitle = 'Chutes x gols marcados')
```



```
# Prevendo a quantidade de chutes para um time com 5 gols
df_chutes <- data.frame(gols = c(5))
predict(mod, df_chutes)
```

```
##          1
## 13.39601
```