

Ultimo Trabalho Regressao

Bruno Mesquita dos Santos

2023-11-15

Biblioteca

```
if (!require('readr'))install.packages("readr");library(readr)
```

```
## Loading required package: readr
```

```
if (!require('faraway'))install.packages("faraway");library(faraway)
```

```
## Loading required package: faraway
```

Dados Consumo Cerveja

```
Consumo_cerveja <- read_csv("C:/Users/onurb/Downloads/Consumo_cerveja.csv",  
                             locale = locale(decimal_mark = ","))
```

```
## Rows: 941 Columns: 7
```

```
## -- Column specification -----
```

```
## Delimiter: ","
```

```
## dbl   (5): Temperatura Media (C), Temperatura Minima (C), Temperatura Maxima ...
```

```
## date  (1): Data
```

```
##
```

```
## i Use 'spec()' to retrieve the full column specification for this data.
```

```
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

Funções

```
rm_accent <- function(x) iconv(x, to = "ASCII//TRANSLIT")
```

Limpeza

```
names(Consumo_cerveja) <- tolower(  
  gsub(' ', '_', gsub('[ ]', '', gsub('[()]', '', rm_accent(names(Consumo_cerveja)))))  
)  
Consumo_cerveja <- na.omit(Consumo_cerveja)
```

```
##Separar Bases
```

```
data <- Consumo_cerveja$data
temperatura_media_c <- Consumo_cerveja$temperatura_media_c
temperatura_minima_c <- Consumo_cerveja$temperatura_minima_c
temperatura_maxima_c <- Consumo_cerveja$temperatura_maxima_c
precipitacao_mm <- Consumo_cerveja$precipitacao_mm
final_de_semana <- Consumo_cerveja$final_de_semana
consumo_de_cerveja_litros <- Consumo_cerveja$consumo_de_cerveja_litros
```

Questão A

Ajuste um modelo de regressão linear múltipla considerando todas as variáveis independentes. Verifique a multicolinearidade entre as variáveis independentes, e se há necessidade de excluir alguma delas por esse critério. Em caso afirmativo, ajuste novo modelo sem essa variável. Apresente todos os valores de Vif.

```
modelo_completo = lm(consumo_de_cerveja_litros ~ data+temperatura_media_c+
                     temperatura_minima_c+temperatura_maxima_c+
                     precipitacao_mm+final_de_semana)

summary(modelo_completo)
```

```
##
## Call:
## lm(formula = consumo_de_cerveja_litros ~ data + temperatura_media_c +
##     temperatura_minima_c + temperatura_maxima_c + precipitacao_mm +
##     final_de_semana)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4823.8 -1794.4  -193.9  1804.6  5514.2
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -60845.145   19516.939   -3.118  0.001971 **
## data              4.027     1.167    3.451  0.000625 ***
## temperatura_media_c   -30.772    186.065   -0.165  0.868737
## temperatura_minima_c    46.816    110.380    0.424  0.671725
## temperatura_maxima_c   675.535     93.905    7.194  3.72e-12 ***
## precipitacao_mm     -58.485     9.892   -5.913  7.86e-09 ***
## final_de_semana    5198.875    267.020   19.470  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2298 on 358 degrees of freedom
## Multiple R-squared:  0.7316, Adjusted R-squared:  0.7271
## F-statistic: 162.6 on 6 and 358 DF, p-value: < 2.2e-16
```

```
summary(aov(modelo_completo))
```

```
##              Df    Sum Sq  Mean Sq F value    Pr(>F)
```

```
## data          1 1.335e+07 1.335e+07 2.528 0.113
## temperatura_media_c 1 2.384e+09 2.384e+09 451.367 < 2e-16 ***
## temperatura_minima_c 1 2.540e+08 2.540e+08 48.098 1.90e-11 ***
## temperatura_maxima_c 1 3.247e+08 3.247e+08 61.481 5.18e-14 ***
## precipitacao_mm      1 1.752e+08 1.752e+08 33.169 1.82e-08 ***
## final_de_semana      1 2.002e+09 2.002e+09 379.081 < 2e-16 ***
## Residuals          358 1.891e+09 5.282e+06
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Anotação:

Se o VIF for igual a 1 não há multicolinearidade entre os fatores, mas se o VIF for maior que 1, as preditoras podem estar moderadamente correlacionadas. A saída acima mostra que o VIF para os fatores de publicação e anos são cerca de 1.5, o que indica alguma correlação, mas não o suficiente para se preocupar demais com isso. Um VIF entre 5 e 10 indica alta correlação, o que pode ser problemático. E se o VIF for acima de 10, você pode assumir que os coeficientes de regressão estão mal estimados devido à multicolinearidade.

VIF:

```
vif <- 1/(1-summary(modelo_completo)[["r.squared"]])
paste('vif =',vif)
```

```
## [1] "vif = 3.72548692094354"
```

Com relação ao nosso modelo temos um VIF para o modelo completo de 3.7254 o que quer dizer que ele é moderadamente correlacionadas.

```
vif(modelo_completo)
```

```
##          data  temperatura_media_c temperatura_minima_c
##      1.044990          24.129091          6.706766
## temperatura_maxima_c  precipitacao_mm      final_de_semana
##      11.327898          1.039824          1.003914
```

Porém quando vamos para um VIF mais detalhado temos em `temperatura_media_c`, `temperatura_minima_c`, `temperatura_maxima_c` uma multicolinearidade alto, o que faz sentido devido serem dados de temperaturas tendo a mesmas representatividade, logo removeria dois ficando é com a `temperatura_media_c`.

Removendo a multicolinearidade

```
modelo_completo = lm(consumo_de_cerveja_litros ~ data+temperatura_media_c+
                      precipitacao_mm+final_de_semana)
summary(modelo_completo)
```

```
##
## Call:
## lm(formula = consumo_de_cerveja_litros ~ data + temperatura_media_c +
##      precipitacao_mm + final_de_semana)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6190.9 -1911.3  -251.5   1995.6   6504.4
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -69560.767   21094.827   -3.298  0.001073 **
## data              4.556       1.263    3.608  0.000353 ***
## temperatura_media_c    854.735     41.965   20.368 < 2e-16 ***
## precipitacao_mm     -74.587     10.681   -6.983  1.4e-11 ***
## final_de_semana    5239.630     293.724   17.839 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2530 on 360 degrees of freedom
## Multiple R-squared:  0.673, Adjusted R-squared:  0.6694
## F-statistic: 185.2 on 4 and 360 DF, p-value: < 2.2e-16
```

```
summary(aov(modelo_completo))
```

```
##              Df      Sum Sq  Mean Sq F value    Pr(>F)
## data              1 1.335e+07 1.335e+07   2.087    0.149
## temperatura_media_c  1 2.384e+09 2.384e+09 372.581 < 2e-16 ***
## precipitacao_mm      1 3.074e+08 3.074e+08  48.038 1.94e-11 ***
## final_de_semana      1 2.036e+09 2.036e+09 318.216 < 2e-16 ***
## Residuals          360 2.303e+09 6.399e+06
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusão:

```
vif <- 1/(1-summary(modelo_completo)[["r.squared"]])
paste('vif =',vif)
```

```
## [1] "vif = 3.05811655234158"
```

Com relação ao nosso modelo temos um VIF para o modelo completo de 3.058 o que quer dizer que ele é moderadamente correlacionadas.

```
vif(modelo_completo)
```

```
##              data temperatura_media_c    precipitacao_mm    final_de_semana
##              1.010045              1.013175              1.000704              1.002719
```

Porém quando vamos para um VIF mais detalhado temos que não há multicolinearidade entre os fatores.

Questão B

Escreva as hipóteses, decisão e conclusão do teste F para o modelo. Use o pvalor da saída do software para o teste. Faça a interpretação do coeficiente de determinação.

```
summary(modelo_completo)
```

```
##
## Call:
## lm(formula = consumo_de_cerveja_litros ~ data + temperatura_media_c +
##      precipitacao_mm + final_de_semana)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6190.9 -1911.3  -251.5   1995.6  6504.4
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -69560.767   21094.827   -3.298  0.001073 **
## data           4.556       1.263    3.608  0.000353 ***
## temperatura_media_c    854.735     41.965   20.368 < 2e-16 ***
## precipitacao_mm    -74.587     10.681   -6.983  1.4e-11 ***
## final_de_semana    5239.630     293.724   17.839 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2530 on 360 degrees of freedom
## Multiple R-squared:  0.673, Adjusted R-squared:  0.6694
## F-statistic: 185.2 on 4 and 360 DF, p-value: < 2.2e-16
```

Hipótese:

$$\begin{cases} H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0 \\ H_1 : \text{pelo menos um } \beta \text{ difere de zero} \end{cases}$$

Conclusão

Dado que nosso modelo completo possui **p-value: < 2.2e-16**, podemos concluir que há qualquer nível de significância ele rejeita H_0