

# Estratégias Inferenciais

Edgar Luiz de Lima

11/03/2022

## Dados

O conjunto de dados que iremos utilizar foi retirado de um conjunto de dados muito maior de Mazerolle & Desrochers (2005).

Ele apresenta as seguintes variáveis:

1. Shade: variável binária indicando se o indivíduo foi coletado na sombra(1= sim, 0= não).
2. Cloud: cobertura de nuvem expressa em porcentagem.
3. Airtemp: temperatura do ar em graus celcius.
4. Mass\_lost: perda de massa corporal em gramas.
5. Aqui iremos definir o diretório, carregar os pacotes e o banco de dados.

```
library(MuMIn)
library(openxlsx)
dados<- read.xlsx("Dados.xlsx")
```

1.2. Vamos visualizar as 10 primeiras linhas do conjunto de dados.

```
head(dados, 10)
```

```
##      Shade Cloud Airtemp Mass_lost
## 1         0    20      31      8.3
## 2         0    20      31      3.6
## 3         0    20      31      4.7
## 4         0     5     22      7.0
## 5         0     5     22      7.7
## 6         0     5     22      1.6
## 7         0    10     25      6.4
## 8         0    10     25      5.9
## 9         0    10     25      2.8
## 10        0     5     23      3.4
```

1.3. Agora precisamos checar como o formato em que o R leu os dados, para saber se os dados estão no formato correto.

```
str(dados)
```

```
## 'data.frame':   121 obs. of  4 variables:
## $ Shade      : num  0 0 0 0 0 0 0 0 0 0 ...
## $ Cloud       : num  20 20 20 5 5 5 10 10 10 5 ...
## $ Airtemp     : num  31 31 31 22 22 22 25 25 25 23 ...
## $ Mass_lost   : num  8.3 3.6 4.7 7 7.7 1.6 6.4 5.9 2.8 3.4 ...
```

Podemos ver que a variável Shade está interpretada como variável numérica, mas na verdade ela deve ser interpretada como uma variável categórica.

Então vamos converter ela em fator, e depois checar se ela realmente foi interpretada como fator.

```
dados$Shade <- as.factor(dados$Shade)
str(dados)
```

```
## 'data.frame': 121 obs. of 4 variables:
## $ Shade : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ Cloud : num 20 20 20 5 5 5 10 10 10 5 ...
## $ Airtemp : num 31 31 31 22 22 22 25 25 25 23 ...
## $ Mass_lost: num 8.3 3.6 4.7 7 7.7 1.6 6.4 5.9 2.8 3.4 ...
```

Agora o R está interpretando ela como fator e nos dizendo que ela possui os níveis 0 e 1 que significa não estar na sombra e estar na sombra respectivamente.

## Testando as nossas hipóteses

Antes de testar as nossas hipóteses precisamos entender primeiro como descrevemos o nosso modelo estatístico no R. Para isso utilizamos o a função `lm()` que significa linear model, e especificamos o nosso modelo da seguinte maneira.

```
lm(y~x, data)
```

onde: `y=` é a variável resposta, nesse caso `Mass_lost`; `x=` é a variável preditora, no nosso caso teremos três variáveis preditoras; `data=` é o conjunto de dados que as nossas variáveis fazem parte. `til(~)=` significa em função de, ou seja, minha variável resposta em função das variáveis preditoras.

### 1. Teste de significância da hipótese nula (NHST)

1.1 Primeiramente, temos que especificar o nosso modelo. Então, especificaremos o nosso modelo da seguinte maneira, `Mass_lost` em função de `Airtemp`, `Cloud` e `Shade`. E indicamos que as variáveis estão dentro do conjunto de dados chamado `dados`.

```
nhst <- lm(Mass_lost~Airtemp+Cloud+Shade, data = dados)
nhst

##
## Call:
## lm(formula = Mass_lost ~ Airtemp + Cloud + Shade, data = dados)
##
## Coefficients:
## (Intercept)      Airtemp        Cloud        Shade1
##      2.420982     -0.005843     -0.006114     -0.789149
```

Pronto, o modelo já foi ajustado, podemos ler o ajuste do modelo da seguinte forma:

1. Intercept= média da `Mass_loss` para os indivíduos que não estão na sombra.
2. `Airtemp=` é o tamanho de efeito(a1) da `Airtemp` sobre a `Mass_loss` (aqui o tamanho de efeito é representado pelo coeficiente de inclinação da reta).
3. `Cloud=` é o tamanho de efeito(a1) da `Cloud` sobre a `Mass_loss`.
4. `Shade1=` é o tamanho de efeito da variável `Shade` sobre o `Mass_loss`, ou seja, é a diferença entre as médias de `Mass_loss` no tratamento 1 e 0.

Podemos obter o valor da média de `Mass_loss` no tratamento `shade1` subtraindo do intercept o valor do tamanho de efeito.

```
coefi<- nhst$coefficients
coefi[1]-abs(coefi[4])
```

```
## (Intercept)
##      1.631833
```

Podemos observar então que o valor da média de Mass\_loss em Shade1 que são aqueles indivíduos coletados em ambiente sombreado é igual a 1.631833

1.2. Agora vamos verificar o valor de p para saber se iremos rejeitar a hipótese nula.

```
summary(nhst)
```

```
##
## Call:
## lm(formula = Mass_lost ~ Airtemp + Cloud + Shade, data = dados)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.2807 -1.0398 -0.5520  0.7802  6.1824
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.420982   1.020847   2.372   0.0193 *
## Airtemp     -0.005843   0.035398  -0.165   0.8692
## Cloud       -0.006114   0.005177  -1.181   0.2401
## Shade1      -0.789149   0.328119  -2.405   0.0177 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.596 on 117 degrees of freedom
## Multiple R-squared:  0.07631,    Adjusted R-squared:  0.05263
## F-statistic: 3.222 on 3 and 117 DF,  p-value: 0.02524
```

Intervalo de confiança

```
intconf<- lm(Mass_lost~Airtemp+Cloud+Shade, data = dados)
confint.lm(object = intconf)
```

```
##              2.5 %      97.5 %
## (Intercept)  0.3992479  4.442716366
## Airtemp     -0.0759474  0.064260803
## Cloud       -0.0163671  0.004139941
## Shade1      -1.4389721 -0.139325345
```

```
coef(intconf)
```

```
## (Intercept)      Airtemp      Cloud      Shade1
##  2.420982139 -0.005843298 -0.006113580 -0.789148745
```

Selecao de modelos

```
comp<-lm(Mass_lost~Airtemp+Cloud+Shade, data = dados)
tempcl<-lm(Mass_lost~Airtemp+Cloud, data = dados)
tempsh<-lm(Mass_lost~Airtemp+Shade, data = dados)
```

```
contraste<- model.sel(comp,tempcl, tempsh)
contraste<-contraste[,c(6:10)]
```

```
## Warning in '[.model.selection'(contraste, , c(6:10)): cannot recalculate
## "weights" on an incomplete object
```

```
contraste
```

	df	logLik	AICc	delta	weight
## tempsh	4	-226.9641	462.2731	0.0000000	0.55553752
## comp	5	-226.2474	463.0165	0.7434163	0.38307393
## tempcl	4	-229.1668	466.6785	4.4054254	0.06138856