

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)
library(ggplot2)
library(ggpubr)
library(effectsize)
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 sendo 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 está nos dizendo que ela possui os níveis 0 e 1, que significam 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 a função `lm()` que significa linear model, e especificamos o nosso modelo da seguinte maneira.

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

onde: 1. `y` = é a variável resposta, nesse caso `Mass_lost`; 2. `x` = é a variável preditora, no nosso caso teremos três variáveis preditoras; 3. `data` = é o conjunto de dados que as minhas variáveis fazem parte. 4. `til(~)` = significa em função de, ou seja, minha variável resposta em função das variáveis preditoras.

1. Teste de significancia 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 chamada `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 (a_1) 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 (a_1) 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
```

Nesse caso então, rejeitamos a hipótese nula e aceitamos a hipótese alternativa, e de acordo com os nossos resultados, há diferença na perda de massa corporal entre indivíduos que estão na sombra e indivíduos que não estão na sombra. Sendo que indivíduos que não estão na sombra perdem 0.789 gramas a mais que os indivíduos que estão na sombra.

2. Intervalo de confiança

```
intconf<- lm(Mass_lost~Airtemp+Cloud+Shade, data = dados)
ic<-confint.lm(object = intconf)
tam<- coef(intconf)
esnat<- data.frame(c("Air Temp","Cloud","Shade"),tam[c(2,3,4)],ic[c(2,3,4),1],ic[c(2,3,4),2])
colnames(esnat)<- c("Var","Tam", "ICmin", "ICmax")
esnat[,1]<-as.factor(esnat[,1])
```

Aqui a melhor maneira de se conferir os resultados é a partir da visualização gráfica, então vamos construir um gráfico para visualizarmos os resultados.

Mas uma coisa importante de se considerar, é que cada variável está em uma unidade de medida diferente, se vamos analisar os efeitos separadamente, sem fazer comparação entre variáveis, podemos manter os tamanhos de efeito na escala original das variáveis, se formos fazer uma comparação do tamanho de efeito entre as

variáveis precisamos padronizar os tamanhos de efeito. Isso pode ser feito de duas formas, padronizando os tamanhos de efeito depois que o modelo foi ajustado, ou padronizar as variáveis (e.g. Z-Score) antes e depois ajustar o modelo com as variáveis padronizadas.

Aqui iremos fazer a padronização com o modelo já ajustado, a transformação que será aplicada é a do Z-Score.

```
efpad<- effectsize(intconf)

efpad2<- data.frame(c("Air Temp","Cloud","Shade" ),
                    efpad[c(2,3,4),2], efpad$CI_low[-1],efpad$CI_high[-1] )
colnames(efpad2)<- c("Var","Tam", "ICmin","ICmax")
efpad2[,1]<-as.factor(efpad2[,1])
```

Agora vamos criar os gráficos para interpretar os resultados.

Gráfico para o tamanho de efeito na escala original da variável.

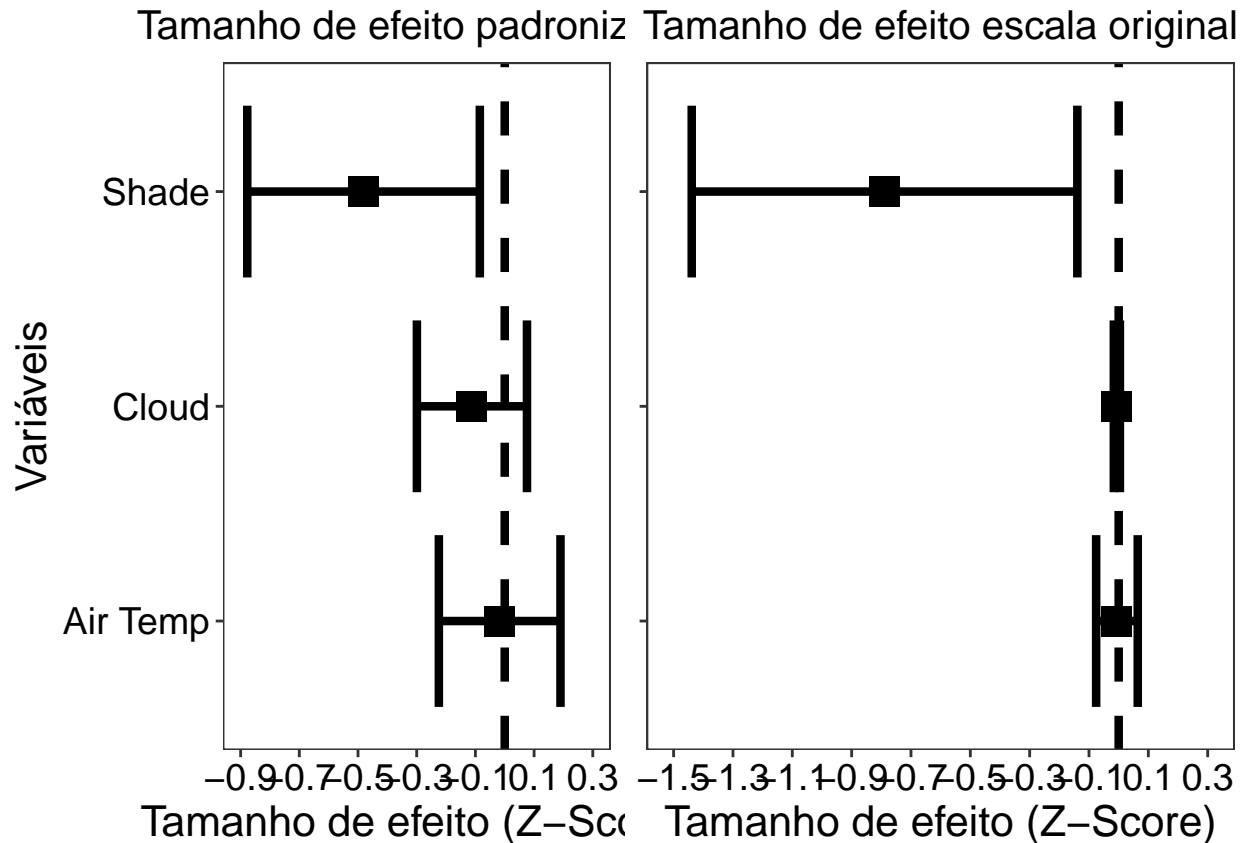
```
orig<-ggplot(data= esnat, aes( Tam,Var,xmin= ICmin, xmax= ICmax))+
  geom_errorbarh(alpha=2,size= 1.5, height= 0.8) +
  geom_point(shape= 15, size=5)+
  scale_x_continuous(breaks= seq(-1.5,0.3,0.2),
                    limits = c(-1.5,0.3))+
  geom_vline(xintercept = 0.0, linetype=2, size = 1.5)+
  ylab("Variáveis")+
  xlab("Tamanho de efeito (Z-Score)")+
  labs(title= "Tamanho de efeito escala original")+
  theme_bw()+
  theme(panel.grid = element_blank(), axis.title = element_text(size= 16, colour = "black"),
        axis.title.y = element_blank(),axis.text = element_text(size = 14, colour = "black"), axis.text.x = element_text(size = 14, colour = "black"),
        plot.title = element_text(size = 15, hjust =.5))
```

Gráfico para o tamanho de efeito padronizado.

```
tampad<- ggplot(data= efpad2, aes( Tam,Var,xmin= ICmin, xmax= ICmax))+
  geom_errorbarh(alpha=2,size= 1.5, height= 0.8) +
  geom_point(shape= 15, size=5)+
  scale_x_continuous(breaks= seq(-0.9,0.3,0.2),
                    limits = c(-0.9,0.3))+
  geom_vline(xintercept = 0.0, linetype=2, size = 1.5)+
  ylab("Variáveis")+
  xlab("Tamanho de efeito (Z-Score)")+
  labs(title= "Tamanho de efeito padronizado")+
  theme_bw()+
  theme(panel.grid = element_blank(), axis.title = element_text(size= 16, colour = "black"), axis.text.x = element_text(size = 14, colour = "black"),
        plot.title = element_text(size = 15, hjust =.5))
```

Juntando os dois gráficos.

```
ggarrange(tampad,orig,ncol = 2, nrow = 1)
```



Com base nos resultados na inferência por intervalo de confiança, podemos concluir que apenas a variável Shade possui efeito sob a perda de massa corporal. Sendo que indivíduos que estão na sombra perdem 0.789 a menos que indivíduos fora da sombra.

3. Selecao de modelos

Para a seleção de modelos iremos testar as seguintes hipóteses: 1. (H1) A Perda de massa é afetada pela Airtemp, Cloud e Shade; 2. (H2) A perda de massa é afetada pela Airtemp e Cloud; 3. (H3) A perda de massa é afetada pela Airtemp e Shade; 4. (H4) A perda de massa é afetada pela Airtemp.

Ajustando os modelos correspondentes a cada uma das hipóteses.

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

3.1. Seleção de modelos utilizando a Soma dos quadrados dos resíduos (SQR).

```
sqr<-data.frame(sum(comp$residuals^2),sum(tempcl$residuals^2), sum(tempsh$residuals^2),sum(temp$residuals^2))
colnames(sqr)<- c("H1", "H2", "H3", "H4")
sqr
```

```
##          H1          H2          H3          H4
## 1 298.133 312.8723 301.686 301.686
```

3.2. Seleção de modelos utilizando o R^2 ajustado.

```

a<- summary(comp)$adj.r.squared
b<- summary(tempcl)$adj.r.squared
c<- summary(tempsh)$adj.r.squared
d<- summary(temp)$adj.r.squared

raju<-data.frame(a,b,c,d)
colnames(raju)<- c("H1", "H2", "H3", "H4")
raju

```

```

##           H1           H2           H3           H4
## 1 0.05262928 0.0142179 0.04946321 0.04946321

```

3.3. Seleção de modelos utilizando o AIC.

```

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

```

```

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

```

```

contraste[,c(1,3)]

```

```

##      df      AICc
## tempsh 4 462.2731
## temp   4 462.2731
## comp   5 463.0165
## tempcl 4 466.6785

```

Organizando os dados para fazer a análise gráfica.

```

daic<- data.frame(row.names(contraste),contraste[,4])
colnames(daic)<- c("Hipoteses", "Delta_AIC")

```

Seleção de modelos utilizando o Delta AIC e fazendo a interpretação gráfica.

```

ggplot(data= daic, aes( Hipoteses,Delta_AIC))+
  geom_point(shape= 15, size=5)+
  #scale_x_continuous(breaks= seq(-0.9,0.3,0.2),
  # limits = c(-0.9,0.3))+
  geom_hline(yintercept = 2, linetype=2, size = 1.5)+
  ylab(~Delta~AIC)+
  xlab("Hipóteses")+
  #labs(title= "Tamanho de efeito padronizado")+
  theme_bw()+
  theme(panel.grid = element_blank(), axis.title = element_text(size= 16, colour = "black"), axis.text =
    plot.title = element_text(size = 15, hjust =.5))

```

