

Lista 2 Computacional

Davi Wentrück Feijó - 200016806

2023-06-19

Questao 1) Considere o modelo

$$Y_{ij} \sim \text{Poisson}(\lambda_{ij})$$
$$\lambda_{ij} = \exp(\alpha + \beta t_i),$$

onde Y_{ij} representa o número de montanhistas que atingiram o cume pela rota j no dia i , α e β são parâmetros desconhecidos do modelo e t_i indica a temperatura média no dia i . Para uma determinada temperatura, o modelo prevê o mesmo valor para todas as rotas. Desconsidere os dados da rota *glacier only - no summit attempt*. Por fim, note que para obter o valor de y_{ij} é preciso somar todos os sucessos registrados no dia i para a rota j .

```
set.seed(420)
#Organizando o banco de dados
climbing <- read_csv("climbing_statistics.csv")
weather <- read_csv("Rainier_Weather.csv")
convert <-function(x) (x-32) * 5/9
shift <-function(x) x - mean(x)
dados <- inner_join(climbing, weather) %>%
  select(-matches("Percent|Battery")) %>%
  filter(Attempted >= Succeeded) %>%
  mutate(`Temperature AVG` = convert(`Temperature AVG`),
         Cleaver = Route == "Disappointment Cleaver",
         Date = lubridate::mdy(Date)) %>%
  select(Date, Succeeded, everything()) %>%
  rename(Date = Date,
         Sucessos = Succeeded,
         Tentativas = Attempted,
         Temperatura = `Temperature AVG`,
         Umidade_relativa = `Relative Humidity AVG`,
         Velocidade_vento = `Wind Speed Daily AVG`,
         Direc_vento = `Wind Direction AVG`,
         Radiacao_solar = `Solare Radiation AVG`)

# Agrupando por dia e rota e desconsiderando a rota
dados2 <- dados %>% group_by(Date,Route) %>%
  mutate(Sucessos = sum(Sucessos),
         Tentativas = sum(Tentativas),
         Cleaver = case_when(Cleaver == T ~ 1,
                             T ~ 0)) %>%
  summarise_if(is.numeric,mean) %>%
  filter(Route != "glacier only - no summit attempt")
```

```
head(climbing)
```

```
## # A tibble: 6 x 5
##   Date      Route      Attempted Succeeded 'Success Percentage'
##   <chr>     <chr>         <dbl>     <dbl>         <dbl>
## 1 11/27/2015 Disappointment Cleaver      2         0           0
## 2 11/21/2015 Disappointment Cleaver      3         0           0
## 3 10/15/2015 Disappointment Cleaver      2         0           0
## 4 10/13/2015 Little Tahoma          8         0           0
## 5 10/9/2015  Disappointment Cleaver      2         0           0
## 6 10/3/2015  Disappointment Cleaver     10         0           0
```

```
head(weather)
```

```
## # A tibble: 6 x 7
##   Date      'Battery Voltage AVG' 'Temperature AVG' 'Relative Humidity AVG'
##   <chr>         <dbl>         <dbl>         <dbl>
## 1 12/31/2015      13.8          19.1          21.9
## 2 12/30/2015      13.8          14.6          18.5
## 3 12/29/2015      13.8           6.61         34.1
## 4 12/28/2015      13.7           8.69         70.6
## 5 12/27/2015      13.4          14.1          95.8
## 6 12/26/2015      13.5          17.5          47.6
## # i 3 more variables: 'Wind Speed Daily AVG' <dbl>, 'Wind Direction AVG' <dbl>,
## #   'Solare Radiation AVG' <dbl>
```

```
head(dados)
```

```
## # A tibble: 6 x 10
##   Data      Sucessos Route      Tentativas Temperatura Umidade_relativa
##   <date>     <dbl> <chr>         <dbl>         <dbl>         <dbl>
## 1 2015-11-27      0 Disappointment Cl~      2      -3.15          19.7
## 2 2015-11-21      0 Disappointment Cl~      3     -0.389          21.7
## 3 2015-10-15      0 Disappointment Cl~      2       8.03          27.2
## 4 2015-10-13      0 Little Tahoma          8       4.99          28.3
## 5 2015-10-09      0 Disappointment Cl~      2       3.48          74.3
## 6 2015-10-03      0 Disappointment Cl~     10     -0.0984          62.3
## # i 4 more variables: Velocidade_vento <dbl>, Direc_vento <dbl>,
## #   Radiacao_solar <dbl>, Cleaver <lgl>
```

```
head(dados2)
```

```
## # A tibble: 6 x 10
## # Groups:   Data [6]
##   Data      Route      Sucessos Tentativas Temperatura Umidade_relativa
##   <date>     <chr>         <dbl>     <dbl>         <dbl>         <dbl>
## 1 2014-09-23 Disappointment Cl~      0         11       0.476          100
## 2 2014-09-24 Disappointment Cl~      0         12      -1.28          100
## 3 2014-09-25 Disappointment Cl~      8         14      -2.88          99.9
## 4 2014-09-26 Disappointment Cl~     19         41      -2.09          100
## 5 2014-09-27 Disappointment Cl~     25         44       0.733          60.2
```

```
## 6 2014-09-28 Disappointment Cl~      6      23      6.13      12.4
## # i 4 more variables: Velocidade_vento <dbl>, Direc_vento <dbl>,
## #   Radiacao_solar <dbl>, Cleaver <dbl>
```

```

# Número de simulações
n <- 10000

# Arrumando a variável Rota
dados3 <- dados2 %>% mutate(Rota = case_when(Route == "Disappointment Cleaver"~"Disappointment Cleaver"
  select(Sucessos, Rota) %>%
  arrange(Rota)

# Calculando a diferença absoluta entre as médias dos grupos

diff_media = abs(as.numeric(diff(tapply(dados3$Sucessos,dados3$Rota,mean))))

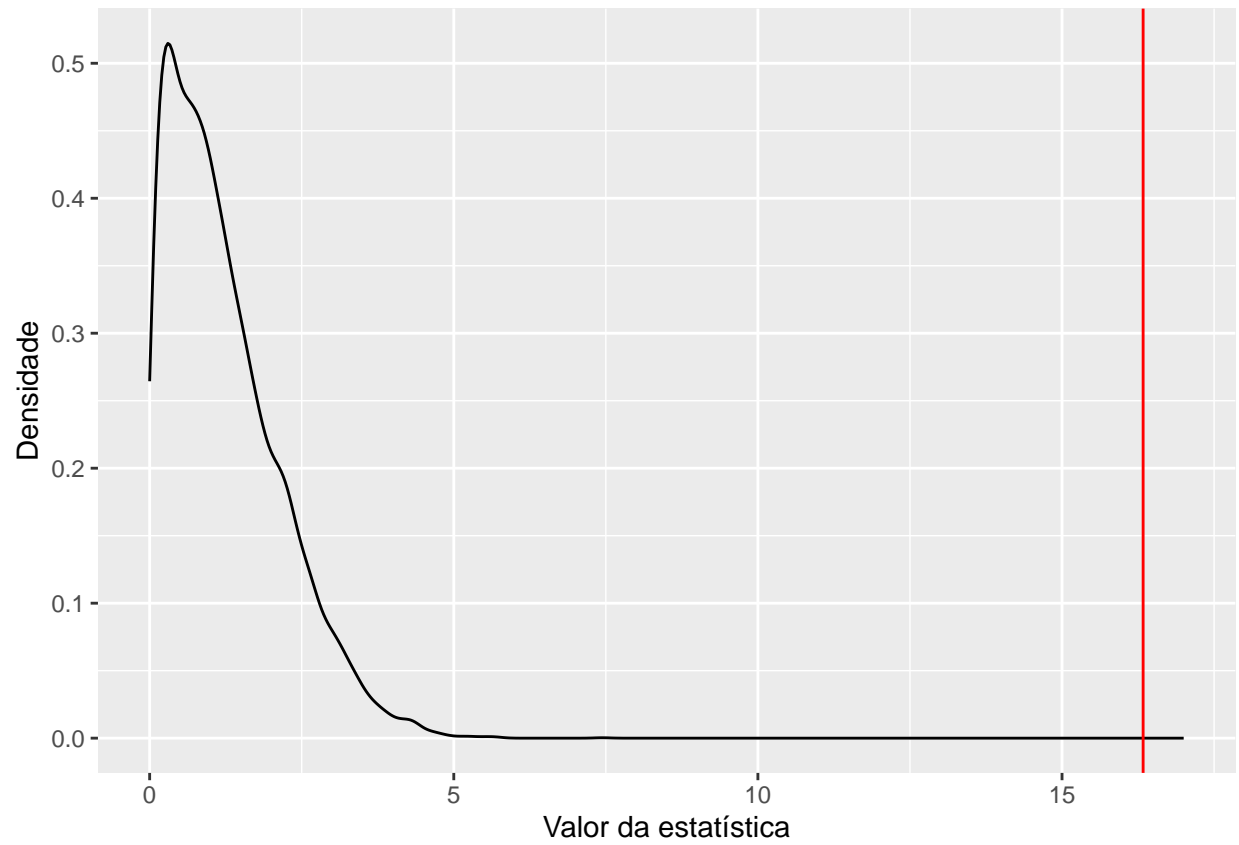
x <- lapply(1:n, function(a) {
  dados_sinteticos <- dados3
  dados_sinteticos$Sucessos <- sample(dados3$Sucessos, replace = FALSE)
  diff_media_calculada <- abs(as.numeric(diff(tapply(dados_sinteticos$Sucessos,dados_sinteticos$Rota,me
  diff_media_calculada
}))

results <- data.frame(estatistica = unlist(x))

# Gráfico da distribuição da diferença entre as médias e a posição da diferença entre a média real em v
ggplot(data = results, aes(x = estatistica)) +
  geom_density() +
  geom_vline(xintercept = diff_media,
    color = "red") +
  xlim(0, 17) +
  xlab("Valor da estatística") +
  ylab("Densidade")

```

a) Conduza um teste de hipóteses por simulação para avaliar a hipótese nula de que a média do número de sucessos obtidos pela rota “Disappointment Cleaver” é igual a média das demais rotas (conjuntamente).



Podemos perceber que o valor é observado é bem extremo, logo podemos assumir que existe uma diferença de médias entre os grupos

```

# Variáveis
intercepto <- 1
temperatura <- dados2$Temperatura
sucessos <- dados2$Sucessos

logistica <- function(eta) exp(eta)
# Função para estimação dos parâmetros
log_verossimilhanca <- function(parametros, t, y) {
  eta <- t %*% parametros
  lambda <- logistica(eta)
  soma <- sum(dpois(y, lambda, log = TRUE))
  return(soma)
}

parametros_iniciais <- c(0, 0)
ajuste <- optim(parametros_iniciais, function(p) -log_verossimilhanca(p, t = cbind(intercepto, temperatura), y = sucessos),
               hessian = TRUE)
# Parâmetros estimados
parametros_est <- ajuste$par
parametros_est

```

b) Obtenha o estimador de máxima verossimilhança de α e β considerando o modelo proposto. Dica: Use a função `optim` do R para achar o ponto que maximiza a log-verossimilhança.

```
## [1] 1.98571963 0.08171069
```

```

# Modelo linear generalizado
mod <- glm(Sucessos ~ Temperatura, data = dados2, family=poisson())
tidy(mod)

```

```
## # A tibble: 2 x 5
##   term          estimate std.error statistic   p.value
##   <chr>          <dbl>    <dbl>    <dbl>   <dbl>
## 1 (Intercept)    1.99      0.0210     94.4 0
## 2 Temperatura    0.0817    0.00283    28.9 1.85e-183
```

Comparando os resultados podemos observar que os parametros estao bem proximos do esperado

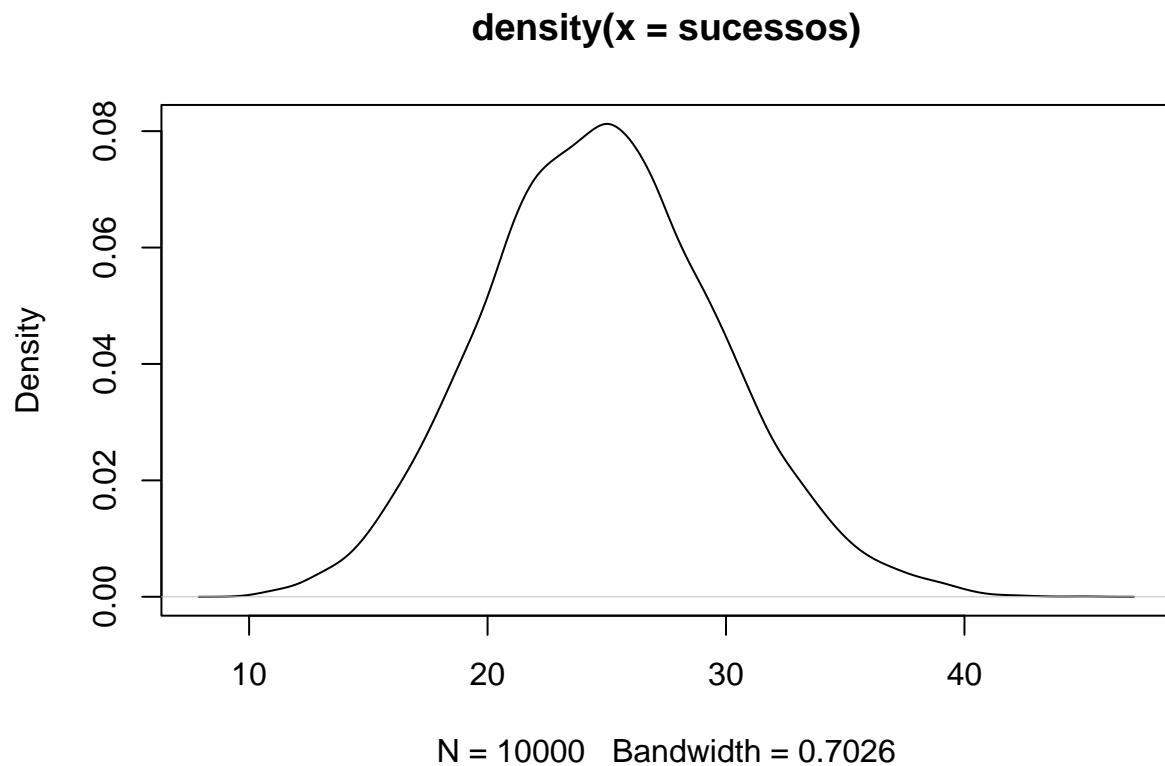
```

# alpha e beta estimados
a <- parametros_est[1]
b <- parametros_est[2]
# Distribuição de n = 1000 observações geradas a partir do alpha e beta estimados para uma temperatura
sucessos <- rpois(n,exp(a + b * 15))

# Distribuição dos sucessos gerados
plot(density(sucessos))

```

c) Estime a distribuição de probabilidade do número de sucessos previstos para um dia em que a temperatura seja de 15 graus.



```

# Número de simulações
m <- 1000

# Variáveis
t <- dados2$Temperatura
y <- dados2$Sucessos
n <- nrow(dados2)

# Simulação
results <- unlist(lapply(1:m, function(amostra) {
  lambda <- exp(a + b * t)
  y.sim <- rpois(n, lambda)
  lambda2 <- exp(a + b * (t + 1))
  y.sim2 <- rpois(n, lambda2)
  log(mean(y.sim2) / mean(y.sim))
}))

# Intervalo de confiança para Beta com 5% de confiança
ic <- quantile(results, probs = c(0.025, 0.5, 0.975))
ic

```

d) Construa um intervalo de confiança de 95% para $\exp(\beta)$ a partir do método de bootstrap paramétrico. Interprete o resultado considerando o contexto dos dados. Dica: calcule o aumento percentual da média esperada quando a temperatura aumenta em 1 grau Celsio.

```

##          2.5%          50%          97.5%
## 0.04457929 0.08294272 0.12120441

```

Como o beta ela representa a taxa de alteracao quando se aumenta a temperatura em uma unidade. Na nossa simulacao podemos encontrar esse valor e estimar uma intervalo de confianca para esse paramentro comparando as taxas de aumento.


```

# Ajuste do modelo original
mod <- lm(Sucessos ~ Temperatura, data = dados2, family = poisson())
t <- dados2$Temperatura
y.est <- fitted(mod)
sigma.est <- summary(mod)$sigma

# Parâmetros do modelo original
a <- coef(mod)[1]
b <- coef(mod)[2]

# Número de repetições e vetor para armazenar os MSEs
num_repeticoes <- 10000
mse_simulacao <- numeric(num_repeticoes)

# Função para simulação e cálculo dos MSEs
simular_e_calcular_mse <- function(i) {
  # Geração de dados sintéticos
  lambda <- exp(a + b * t)
  y.sim <- rpois(length(t), lambda)

  # Ajuste do modelo sobre os dados sintéticos
  mod_sintetico <- lm(y.sim ~ t, family = poisson())
  y.est_sintetico <- fitted(mod_sintetico)

  # Cálculo do MSE
  mean((y.sim - y.est_sintetico)^2)
}

# Simulação e cálculo dos MSEs usando lapply
mse_simulacao <- unlist(lapply(1:num_repeticoes, simular_e_calcular_mse))

# Cálculo do MSE original
mse_original <- mean((dados2$Sucessos - y.est)^2)

# Cálculo do MSE médio simulado
mse_simulacao_media <- mean(mse_simulacao)

# Comparação e comentários dos resultados
if (mse_simulacao_media > mse_original) {
  mensagem <- "Os MSEs simulados são maiores que o MSE original."
} else if (mse_simulacao_media < mse_original) {
  mensagem <- "Os MSEs simulados são menores que o MSE original."
} else {
  mensagem <- "Os MSEs simulados são iguais ao MSE original."
}

cat("MSE Original:", mse_original, "\n")

```

e) Faça um diagnóstico do modelo via simulação. Para tanto, gere dados sintéticos usando o modelo obtido no item b), ajuste um novo modelo sobre os dados sintéticos e calcule o Erro quadrático médio (MSE). Repita esse procedimento 10000 vezes e compare os MSEs gerados com aquele do modelo obtido em b). Comente os resultados.

```
## MSE Original: 235.4943
```

```
cat("MSE Médio Simulado:", mse_simulacao_media, "\n")
```

```
## MSE Médio Simulado: 1.872729e+14
```

```
cat(mensagem)
```

```
## Os MSEs simulados são maiores que o MSE original.
```

Podemos perceber que o MSE dos dados sintéticos são bem maiores que o MSE dos dados originais.

Questao 2) Use o método de integração por Monte Carlo para estimar o volume de uma elipsoide definida por

$$\frac{x^2}{2} + \frac{y^2}{3} + \frac{z^2}{4} = 1.$$

```
# Número de pontos simulados
```

```
m <- 100000
```

```
# Valor exato
```

```
a <- sqrt(2)
```

```
b <- sqrt(3)
```

```
c <- sqrt(4)
```

```
(v <- (4/3) * pi * a * b * c)
```

```
## [1] 20.5208
```

```
# Simulação 3 dimensões
```

```
x <- runif(m, -a, a)
```

```
y <- runif(m, -b, b)
```

```
z <- runif(m, -c, c)
```

```
# Valores dentro da região do elipsoide
```

```
dentro <- x^2/2 + y^2/3 + z^2/4 < 1
```

```
8 * a * b * c * mean(dentro)
```

```
## [1] 20.56043
```