LISTA 3

ESTATÍSTICA COMPUTACIONAL

Tailine J. S. Nonato

2023-11-30

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, Route != "glacier only - no summit attempt",Route != "U
    mutate(`Temperature AVG`= convert(`Temperature AVG`),Cleaver = Route=="Disappointment
    select(Date, Succeeded, everything()) %>%
    rename(Data = Date,
           Rota = Route,
           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`) %>%
    group_by(Data, Rota) %>%
    mutate(Sucessos = sum(Sucessos), Tentativas = sum(Tentativas)) %>%
    distinct()
kable(head(dados),align='l')
```

Data	Sucessos	Rota	Tentativas	Temperatura	Umidade_relativa	Velocidade_vento	Direcao_vento	Radiacao_solar	Cleaver
2015-11-27	0	Disappointment Cleaver	2	-3.1546296	19.71500	27.839583	68.00417	88.49625	TRUE
2015-11-21	0	Disappointment Cleaver	3	-0.3888889	21.69071	2.245833	117.54967	93.66042	TRUE
2015-10-15	0	Disappointment Cleaver	2	8.0266204	27.21125	17.163625	259.12138	138.38700	TRUE
2015-10-13	0	Little Tahoma	8	4.9886574	28.33571	19.591167	279.77917	176.38267	FALSE
2015-10-09	0	Disappointment Cleaver	2	3.4780093	74.32917	65.138333	264.68750	27.79129	TRUE
2015-10-03	0	Disappointment Cleaver	12	-0.0983796	62.33708	13.125042	153.93167	196.37521	TRUE

Item 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).

- **Respostas**

Tem-se como hipóteses:

$$\begin{cases} H_0 : \mu_1 = \mu_{2,\ldots,n} \\ H_1 : \mu_1 \neq \mu_{2,\ldots,n} \end{cases}$$

Observando primeiramente de forma 'exploratória', calculamos as médias e a diferenças entre elas.

```
tab1<- tidy(summary(dados$Sucessos))
kable(tab1,align='l')
```

Table 2: Sumário de Sucessos

minimum	q1	median	mean	q3	maximum
0	0	2	10.40343	17	92

```r
tab2 <- dados %>% filter(Rota == 'Disappointment Cleaver')
mu1 <- mean(tab2$Sucessos)

tab3 <- dados %>% filter(Rota != 'Disappointment Cleaver')
mu2 <- mean(tab3$Sucessos)
diff <- mu1-mu2

tab4 <- data.frame(mu1,mu2,diff)
kable(tab4,align='l')
```

Table 3: Análitica das Médias

mu1	mu2	diff
20.48876	4.170139	16.31863

Considerando o Súmario de Sucesso, observa-se uma média de em torno de 10.4 sucessos em todas as rotas avaliadas, assim o valor da diferença 16.32 pode ser considerado alto, indicando a possível diferença de médias.

Realizando uma simulação com $n = 10.000$, tem-se que:

```r
otherR <- dados %>%
    filter(Rota != 'Disappointment Cleaver') %>%
    select(Sucessos)
DCR <- dados %>%
filter(Rota == 'Disappointment Cleaver') %>%
    select(Sucessos)

n <- 10000
diff2 <- numeric(n)
for (i in 1:n){
    mu11 <- sample(DCR$Sucessos,size=10,replace=T)
    mu22 <- sample(otherR$Sucessos,size=10,replace=T)
    mu11 <- mean(mu11)
    mu22 <- mean(mu22)
    diff2[i] <- abs(mu11 - mu22)
}

kable(tidy(summary(diff2)),align='l')
```

Table 4: Sumário de Diferenças na Simulação

minimum	q1	median	mean	q3	maximum
0	11.5	16.1	16.32942	20.9	45.2

```
prob <- data.frame(paste0(round(mean(diff2)/n*100,2),'%'))
colnames(prob)<-c("Probabilidade")
kable(prob,align='c')
```

Table 5: P(H_0 é aceita | H_0 é verdadeira)

Probabilidade
0.16%

Com uma probabilidade ~0.01 de que as médias seriam iguais, há evidências para rejeitar H_0.

Item 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.

- **Respostas**

```
lv <- function(p,temp,sucesso){
    l <- exp(p[1]+p[2]*temp)
    sum(dpois(sucesso,l,log=T))}
estim <- optim(c(0,0), function(.) -lv(.,temp=dados$Temperatura, sucesso=dados$Sucessos))$

estim<- data.frame(t(estim))
colnames(estim) <- c('Alpha','Beta')
kable(estim,align='c')
```

Table 6: Estimadores de Máxima Verossimilhança

Alpha	Beta
1.988505	0.0812769

Item C

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

- **Respostas**

Para verificar, basta utilizar a mesma estrutura anterior, mas considerando o valor fixo $T_i = 15$.

```
estim2 <- optim(c(0,0), function(.) -lv(.,temp=15, sucesso=dados$Sucessos))$`par`
estim2 <- data.frame(t(estim2))

colnames(estim2) <- c('Alpha','Beta')
kable(estim2,align='c')
```

Table 7: Estimadores de Máxima Verossimilhança

Alpha	Beta
0.137087	0.1470131

Verificando um intervalo [0,95] (min-max sucessos), tem-se que:

```
x <- 0:95
pr <- (estim2[1]+estim2[2])*15
lambda<- as.numeric(exp(pr))
fprob <- dpois(x, lambda)

plot(fprob)
```

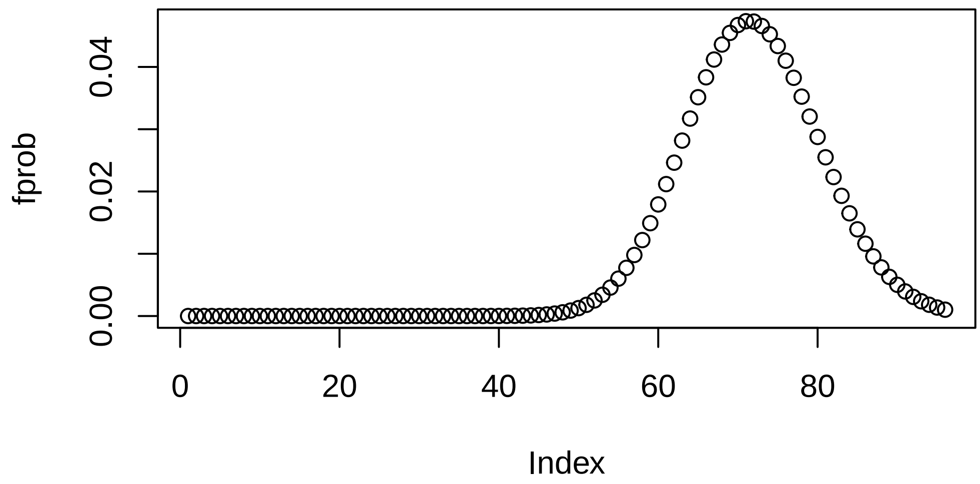

Figure 1: Distribuição de Probabilidade de Sucessos quando Temperatura é 15º C

Item 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.

- **Respostas**

```
sucessos <- numeric()
beta.estim <- numeric()
SE <- numeric()

for(j in 1:1000){
    for(i in 1:nrow(dados)){
        pr <- exp(as.numeric(estim[1]+estim[2]*dados$Temperatura[i]))
        sucessos[i] <- rpois(1,pr)}
    estim3 <- optim(c(0,0), function(.) -lv(., temp=dados$Temperatura, sucesso=sucessos))$
```

```
    beta.estim[j] <- estim3[2]

    SE[j] <- mean((sucessos-exp(as.numeric(estim[1])+as.numeric(estim[2])*dados$Temperatur
    sucessos <- numeric()
    }
y <- exp(beta.estim)
kable(tidy(summary(y)))
```

Table 8: Sumário de $Y \sim \exp(\beta)$

minimum	q1	median	mean	q3	maximum
1.075012	1.082565	1.084684	1.084686	1.086665	1.094451

```
plot(y)
```

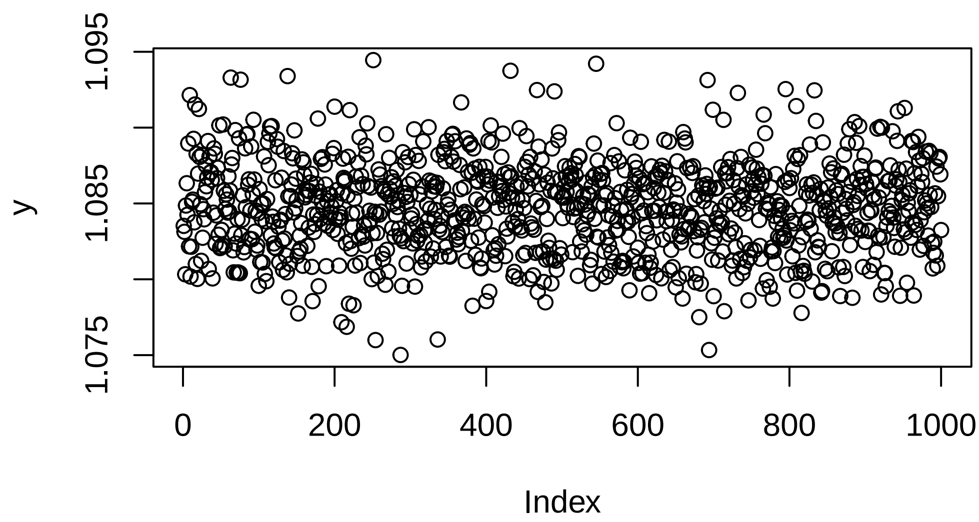

Figure 2: Distribuição de Probabilidade de $Y \sim \exp(\beta)$

```
alpha <- 1-0.95
inf<-as.numeric(quantile(y, probs= (alpha/2)))
sup<-as.numeric(quantile(y, probs= (1-alpha/2)))
```

```
ic<- data.frame(inf,sup)
colnames(ic) <-c('Limite Inferior','Limite Superior')
kable(ic,align='c')
```

Limite Inferior	Limite Superior
1.078998	1.090514

$$IC(\exp(\beta), 95\%) = [1.078, 1.091]$$

Item 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.

- **Respostas**

Com fins de otimização, a simulação foi realizada no Item D, mas apenas com 1000 repetições por limitações de hardware e tempo de execução.

```
MSE <- mean(SE)
kable(MSE,align='c')
```

Table 10: Erro Quadrático Médio

x
0.0215966