Trabalho 1

Daniel Krügel

2022-12-01

Questão 2

Para esta questão vamos adotar como sucesso a presença de clamídia no exame da paciênte, portanto só vamos considerar dois resultados, caso o teste dê positivo e negativo, cada uma das tentativas deve de ser feita em um paciente independente do outro e considerando a proporção de casos positivos constante na população. Com estas suposições será utilizado o modelo binomial para a análise deste caso.

Análise

Crio as variáveis de w e n para esta questão e utilizo o estimador de pi como sendo:

$$\hat{\pi} = w/n$$

```
n <- 750
w <- 48
pi.hat <- w/n

# Y ~ binomial (0.064)
```

Em seguida crio os intervalos de confiança, para este caso como o $\hat{\pi}$ é pequeno e o número da amostra é grande optei por utilizar o intervalo de Agrest-coull, porém criei o intervalo Wald para desencargo de consciência.

```
#Criação do intervalo Wald
alpha <- 0.05
var.wald <- pi.hat*(1 - pi.hat)/n
lower <- pi.hat - qnorm (p = 1- alpha /2) * sqrt ( var.wald )
upper <- pi.hat + qnorm (p = 1- alpha /2) * sqrt ( var.wald )
round ( data.frame (lower , upper ), 4)
```

```
##    lower  upper
## 1 0.0465 0.0815
```

```
#Criação do intervalo Agrest Coull
p.tilde <- (w + qnorm(p = 1- alpha/2)^2/2) / (n + qnorm(p = 1- alpha/2)^2)
var.ac <- p.tilde*(1 -p.tilde ) / (n + qnorm(p = 1- alpha/2)^2)
round(p.tilde + qnorm(p = c( alpha/2, 1- alpha/2) ) * sqrt(var.ac), 4)
```

```
## [1] 0.0485 0.0840
```

A uma probabilidade de 95% da média de permanencia de clamídia entre a população estudada estar entre 0.0840 e 0.0485.

Questão 5

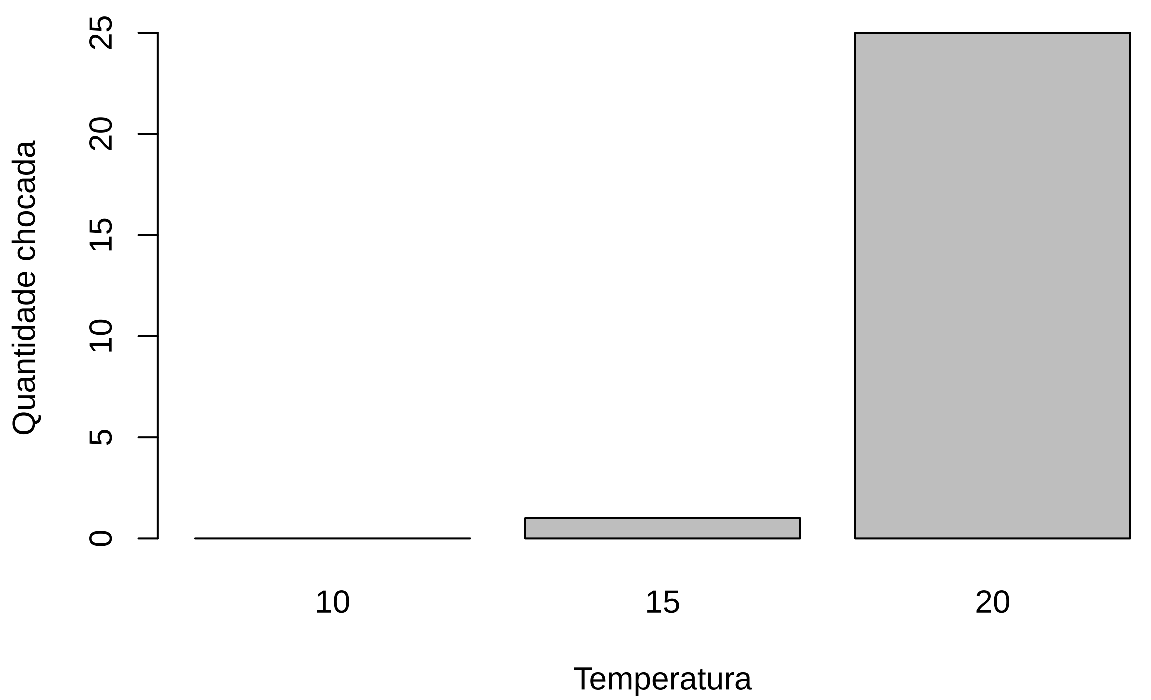
Criando o data frame para o estudo

```
tb <- data.frame(temp = c(10,15,20),
                 qtd = c(0,1,25))
```

Algumas considerações sobre esta questão:

- Temos dados muito diferentes para esta questão, temos estudos que não houve um sucesso sequer
- Me perdoe o erro de não utilizar a temperatura como fator, não vi a necessidade para esta questão mas está categóricamente incorreto.

Para ilustrar a diferença entre a escala de 10 graus e 25 graus:



Posto isto posso acalmar caso encontre algum resultado muito discrepante entre os fatores de temperatura.

Intervalos de confiança

A criação de intervalos de confiança para a proporção estimada de cada uma das temperaturas:

```r
alpha <- 0.05
n <- 30
w <- tb$qtd
pi.hat <- w/n
p.tilde <- (w + qnorm(p = 1-alpha/2)^2/2)/(n+qnorm(1-alpha/2)^2)
# Wald
var.wald<-pi.hat*(1-pi.hat)/n
lower.wald <- pi.hat - qnorm(p = 1-alpha/2)*sqrt(var.wald)
upper.wald <- pi.hat + qnorm(p = 1-alpha/2)*sqrt(var.wald)

Wald.inter <- data.frame(temperatura = tb$temp,
                         lower = lower.wald,
                         upper = upper.wald)
# Agresti-Coull
lower.AC <- p.tilde - qnorm(p = 1-alpha/2) * sqrt(p.tilde*(1-p.tilde) / (n+qnorm(1-alpha/2)^2))
upper.AC <- p.tilde + qnorm(p = 1-alpha/2) * sqrt(p.tilde*(1-p.tilde) / (n+qnorm(1-alpha/2)^2))
AC.inter <- data.frame(temperatura = tb$temp,
                         lower = lower.AC,
                         upper = upper.AC)
# Wilson
lower.wilson <- p.tilde - qnorm(p = 1-alpha/2) * sqrt(n) / (n+qnorm(1-alpha/2)^2) *
  sqrt(pi.hat*(1-pi.hat) + qnorm(1-alpha/2)^2/(4*n))
upper.wilson <- p.tilde + qnorm(p = 1-alpha/2) * sqrt(n) / (n+qnorm(1-alpha/2)^2) *
  sqrt(pi.hat*(1-pi.hat) + qnorm(1-alpha/2)^2/(4*n))

wilson.inter <- data.frame(temperatura = tb$temp,
                         lower = lower.wilson,
                         upper = upper.wilson)

Wald.inter; AC.inter; wilson.inter
```

```
##   temperatura      lower      upper
## 1          10  0.0000000 0.00000000
## 2          15 -0.0309007 0.09756737
## 3          20  0.6999747 0.96669199

##   temperatura        lower     upper
## 1          10 -0.021198303 0.1347117
## 2          15 -0.008305484 0.1809180
## 3          20  0.659603557 0.9313875

##   temperatura      lower     upper
## 1          10 0.00000000 0.1135134
## 2          15 0.00590859 0.1667039
## 3          20 0.66435649 0.9266346
```

Como tivemos uma base de dados pequena ($n = 30$) decidi incluir o intervalo de Wilson, o que se provou uma boa ideia uma vez que foi o unico que se manteve dentro do espaço paramétrico do estimador $\hat{\pi}$.

Teste de hipótese

Conduziremos um teste de hipótese para ver se há diferença na probabilidade de um ovo oclodir em cada uma das temperaturas.

Como a ultima temperatura teve um resultado muito extremo comparado com os outros dois, irei fazer apenas o teste comparando se há diferença entre a chance de um ovo eclodir na temperatura de 20 graus contra o de 15 graus, farei isso utilizando um teste de diferença de proporções, onde a hipótese nula será $\hat{\pi_1} - \hat{\pi_2} = 0$ indicando que não há diferença entre ambas e $\hat{\pi_1} - \hat{\pi_2} > 0$ como a hipótese alternativa, indicando que $\hat{\pi_1}$ é maior que $\hat{\pi_2}$.

```
alpha <- 0.05
pi2 <-  1/30
pi1 <-  5/6
n1 <- 30
n2 <- 30
pi <- 26/60  # w+/n+
#Teste onde pi1 - pi2 == 0 é a hipótese nula e pi1 - pi2 > 0 é a alternativa, portanto é um teste unila

# Estatística teste

Z0 <- (pi1-pi2)/(pi*(1-pi)*((1/n1)+(1/n2))); Z0
```

```
## [1] 48.86878
```

```
Z0 >= qnorm(p = 1-alpha/2)
```

```
## [1] TRUE
```

Valor da estatística teste de 48.86, maior do que o nível de significancia estipulado. Portanto é rejeitada a hipótese nula indicando que $\hat{\pi_1}$ pode ser maior do que $\hat{\pi_2}$ portando $\hat{\pi_3}$ também deve de ser menor.

Questão 13

Criando os dados

```
y <- c(135,434,569,15,9,24,150,443,539)
tb <- matrix(y, nrow = 3, ncol = 3, byrow = TRUE)

alpha <- 0.05
n <- c(tb[1,3],tb[2,3])
w <- c(tb[1,1],tb[2,1])
pi.hat <- w/n
p.tilde <- (w + qnorm(p = 1-alpha/2)^2/2)/(n+qnorm(1-alpha/2)^2)
```

Intervalo de confiança

Para criar um intervalo entre a proporção estimada de cada padrão de uso de preservativo utilizarei um intervalo Wald padrão:

```
var.wald<-pi.hat*(1-pi.hat)/n
lower.wald <- pi.hat - qnorm(p = 1-alpha/2)*sqrt(var.wald)
upper.wald <- pi.hat + qnorm(p = 1-alpha/2)*sqrt(var.wald)
```

```r
data.frame(Padrão = c('Nunca','Sempre'),
                         lower = lower.wald,
                         upper = upper.wald)
```

```
##   Padrão     lower     upper
## 1  Nunca 0.2023048 0.2722119
## 2 Sempre 0.4313141 0.8186859
```

Vemos que a verdadeira chance de se testar positivo para HIV caso nunca utilize preservativo fica em $0.2023 < \hat{\pi}_1 < 0.2722$ e caso sempre utilize fica entre $0.4313 < \hat{\pi}_2 < 0.8186$.

Para ver os intervalos de confiança entre a diferênça entre $\hat{\pi}_1$ e $\hat{\pi}_2$ implementarei as equações em objetos utilizando a função function tanto do intervalo Wald quanto do Agresti-Caffo.

```r
# Criando função para criar intervalo wald para diferença
wald.dif <- function(pi1, pi2, n1 ,n2 , alpha = 0.05){
  var.wald <- (pi1*(1 - pi1)/n1) + (pi2*(1 - pi2)/n2)
  loweR <- pi1-pi2 - qnorm(p = 1-alpha/2) * sqrt(var.wald)
  uppeR <- pi1-pi2 + qnorm(p = 1-alpha/2) * sqrt(var.wald)

  Wald.inter <- data.frame(lower = loweR,
                           upper = uppeR)
  return(Wald.inter)
}

# Criando função para criar intervalo agresti-Caffo para diferença
AgrestCaffo.dif <- function(pi1, pi2, n1 ,n2 , alpha = 0.05){
  var.AC <- (pi1*(1 - pi1)/(n1+2)) + (pi2*(1 - pi2)/(n2+2))
  loweR <- pi1-pi2 - qnorm(p = 1-alpha/2) * sqrt(var.AC)
  uppeR <- pi1-pi2 + qnorm(p = 1-alpha/2) * sqrt(var.AC)

  AC.inter <- data.frame(lower = loweR,
                           upper = uppeR)
  return(AC.inter)
}
```

Utilizando estas funções temos:

```r
pi1 <- tb[1,1]/tb[1,3]
pi2 <- tb[2,1]/tb[2,3]
wald.dif(pi1 = pi1, pi2 = pi2, n2 = tb[1,3], n1 = tb[2,3])
```

```
##        lower      upper
## 1 -0.5625216 -0.2129617
```

```r
AgrestCaffo.dif(pi1 = pi1, pi2 = pi2, n2 = tb[1,3], n1 = tb[2,3])
```

```
##        lower     upper
## 1 -0.5560103 -0.219473
```

Ambos os resultados caem dentro do aspecto negativo indicando que indepente de qual deles seja avaliado $\hat{\pi}_2$ é maior do que $\hat{\pi}_1$ com $(1 - \alpha)\%$ de probabilidade.

Teste de hipótese e razão de chances

Para prosseguir fazendo um teste chi-square e o calculo de odds ratio precisarei criar a tabela de contingência como um array invés de uma matriz.

```
c.table <- array ( data = c(135,15, 434, 9) , dim = c(2 ,2) ,
                   dimnames = list ( First = c("NUNCA", "SEMPRE") , Second = c("Positivo", "Negativo")))
```

```
##         Second
## First    Positivo Negativo
##   NUNCA       135      434
##   SEMPRE       15        9
```

```
prop.test(c.table, conf.level = 0.95, correct = FALSE)
```

```
##
##  2-sample test for equality of proportions without continuity
##  correction
##
## data:  c.table
## X-squared = 18.322, df = 1, p-value = 1.866e-05
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.5845563 -0.1909270
## sample estimates:
##    prop 1    prop 2
## 0.2372583 0.6250000
```

Rejeição de hipõtese nula, existe uma diferença nas médias entre casos positivos de HIV com e sem o uso de preservativo

Valor da estatística Chi quadrado é de 18.322, p valor é de 1.866e-05, dentro da região de rejeição.

```
# Razão de chances

OR <- c.table[1,1] * c.table[2,2]/ (c.table[1,2] * c.table [2,1])

1/OR
```

```
## [1] 5.358025
```

A Não utilização do preservativo incrementa em 5.35 vezes as chances de se testar HIV positivo.

```
alpha<-0.05
var.log.or<-1/c.table[1,1] + 1/c.table[1,2] + 1/c.table[2,1] + 1/c.table[2,2]
OR.CI<-exp(log(OR) + qnorm(p = c(alpha/2, 1-alpha/2)) * sqrt(var.log.or))
round(OR.CI, 2)
```

```
## [1] 0.08 0.44
```

```
1/round(OR.CI, 2)
```

```
## [1] 12.500000  2.272727
```

O resultado do teste Chi quadrado e da razão de chances, tudo indica para que haja uma relação entre o uso de preservativos e a redução de casos testados de HIV positivo.

Questão 17

Criação da tabela de contingência

```
c.table <- array(data = c(118,155,93,51), dim = c(2,2),
                          dimnames = list( First = c("Outra lingua","Nativo"), Second = c("Engraçado","
```

```
##                Second
## First           Engraçado Não
##   Outra lingua        118  93
##   Nativo              155  51
```

```
prop.table(c.table)
```

```
##                Second
## First           Engraçado       Não
##   Outra lingua 0.2829736 0.2230216
##   Nativo       0.3717026 0.1223022
```

```
#Teste chi square
prop.test(c.table)
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c.table
## X-squared = 16.363, df = 1, p-value = 5.229e-05
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.28720861 -0.09916235
## sample estimates:
##    prop 1    prop 2
## 0.5592417 0.7524272
```

P-valor(5,229e-05) caí dentro da área de rejeição de 16,363, indicando que há influência entre encontrar o humor na tira caso a sua primeira língua seja Inglês.

Razão de chances

```r
oddratio <- (c.table[1,1] * c.table[2,2]) / ( c.table[1,2] * c.table [2,1])

1/oddratio
```

```
## [1] 2.395314
```

A razão de chances indica que há um aumento de aproximadamente 2.40 vezes na probabilidade de se encontrar o humor da tira caso a primeira língua seja Inglês.