Profa. Elisângela Rodrigues

Monitoras: Bruna Barreto, Larissa Sousa

Parte II: Prática (usando a linguagem R)

Base de dados `survey` do pacote `MASS` contido no software R.

1) Encontre uma estimativa pontual da idade média do estudante universitário com os dados de amostra da pesquisa `survey`.

Solução:

Em primeiro lugar, chamamos os dados de pesquisa usando `library(MASS)`. Depois, salvamos os dados da pesquisa da idade do aluno. Para encontrar a estimativa pontual da idade média do aluno, escrevemos o código usando a função média. Acontece que nem todos os alunos preencheram a seção de idade, então usamos `na.rm = TRUE` para filtrar o valor ausente.

```
# Apresenta o conjunto de dados survey do pacote MASS
library(MASS)

# Salva a idade do aluno dos dados survey
age_survey = survey$Age

# Encontra a estimativa pontual da idade do aluno
# Acontece que nem todos os alunos preencheram a seção de idade, então devemos filtrar os valores ausentes. Portanto, aplicamos o argume
mean(age_survey, na.rm = TRUE)
```

20.3745147679325

O resultado da codificação da média da estimativa pontual acima é 20,37451 anos.

```
point.estimate <-t.test(age_survey, conf.level = 0.95 )
point.estimate$conf.int
```

19.5459967594874 · 21.2030327763776

Os intervalos de confiança para a idade média do estudante universitário são 19,546 - 21,20303. Portanto, o nível de confiança de 95% inclui a verdadeira média populacional que é igual a 20,37451 anos.

2) Assuma o desvio padrão da população σ da idade do aluno na pesquisa de dados é 7. Encontre a margem de erro e a estimativa de intervalo com nível de confiança de 95%.

Solução:

Precisamos filtrar os valores ausentes em `survey$Age` usando a função `na.omit` e salvá-la como `age.response`. Depois disso, calculamos o erro padrão da média. Portanto, há duas caudas nessa distribuição normal, o nível de confiança de 95% indicaria o percentil 97,5 da distribuição normal na cauda superior. Assim, para obter a margem de erro, multiplicamos a `qnorm(0,975)` pelo erro padrão da média.

```
library(MASS)
age.response = na.omit(survey$Age)
n = length(age.response)

# População de desvio padrão
sigma = 7

# Erro padrão
SE = sigma/sqrt(n)

# Margem de erro
E = qnorm(0.975)*SE
E
```

0.891193392788715

Então descobrimos que a margem de erro é de 0,8911934 anos. Depois disso, somamos com a média amostral para encontrar o intervalo de confiança.

```
# Média da amostra
xbar = mean(age.response)
xbar
```

20.3745147679325

```
# Intervalo de confiança
xbar + c(-E,E)
```

19.4833213751438 · 21.2657081607212

A margem de erro da idade do aluno assumindo que o desvio padrão da população é 7 no nível de confiança de 95% é 0,8911934 anos. O intervalo de confiança para este caso está entre 19,48332 e 21,26571 anos.

Solução alternativa:

Para a solução alternativa, podemos usar a função `z.test` no pacote `TeachingDemos`. Ele também deve ser instalado e carregado no espaço de trabalho.

```
install.packages("TeachingDemos")
library(TeachingDemos)

z.test(age.response, sd=sigma)
```

```
Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)


        One Sample z-test

data:  age.response
z = 44.809, n = 237.0000, Std. Dev. = 7.0000, Std. Dev. of the sample
mean = 0.4547, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 19.48332 21.26571
sample estimates:
mean of age.response
            20.37451
```

3) Sem assumir o desvio padrão da população, σ, da idade do aluno na pesquisa, encontre a margem de erro e a estimativa do intervalo com nível de confiança de 95%.

Solução:

Carregamos primeiro o pacote `MASS` para obter os dados do levantamento. Depois disso, filtramos o valor ausente usando `na.omit`, atribuímos o comprimento usando a função `length` e escrevemos o desvio padrão da amostra. Em seguida, estimamos o erro padrão e a margem de erro (cauda superior de 95% do nível de confiança) como o código abaixo.

```
# Carregar o pacote de MASS
library(MASS)

# Filtra o valor ausente
age.response = na.omit(survey$Age)

# Atribuir o comprimento
n = length(age.response)

# Desvio padrão amostral
s = 7

# Estimando o erro padrão
SE = s/sqrt(n)

# Margem de erro (cauda superior 95% do intervalo de confiança)
E = qt(0.975, df= n -1)*SE
E
```

0.895787155624028

```
# Média da amostra
xbar = mean(age.response)
xbar
```

20.3745147679325

```
# Intervalo de confiança
xbar+c(-E, E)
```

19.4787276123085 · 21.2703019235565

O resultado da margem de erro para a pesquisa de idade do aluno é de 0,8957872 anos com nível de confiança de 95% e o intervalo de confiança está entre 19,47873 e 21,27030 anos.

Solução alternativa:

```
library(stats)
t.test(age.response)
```

```
	One Sample t-test

data:  age.response
t = 48.447, df = 236, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 19.54600 21.20303
sample estimates:
mean of x
 20.37451
```

4) Melhore a qualidade de uma pesquisa amostral aumentando o tamanho da amostra com desvio padrão desconhecido, σ.

Solução:

Como não sabemos qual é o desvio padrão e precisamos melhorar a qualidade de uma pesquisa de amostra, vamos supor que metade do aluno escreva a pesquisa que nos dá a variabilidade máxima. Então, agora o p é 0,5. Agora, digamos que queremos uma margem de erro de 5% e um nível de confiança de 95% que nos dê os valores Z de 1,86.

```
zstar = qnorm(0.975)
(zstar^2*(0.5)* (0.5))/ (0.05)^2
```

384.145882069412

Assim, obtivemos 384,1459 ou 384 tamanhos de amostra para melhorar a qualidade de uma pesquisa amostral com desvio padrão desconhecido σ.

5) Suponha que você não tenha uma estimativa de proporção planejada, encontre o tamanho da amostra necessário para atingir uma margem de erro de 5% para a pesquisa de estudantes do sexo masculino com um nível de confiança de 95%.

Solução:

O que nós sabemos:

- 5% de margem de erro
- Intervalo de confiança de 95%. Então, podemos obter z = 1,96

Primeiro, precisamos descobrir o número de alunos do sexo masculino. Podemos encontrá-lo usando a função `sum` e dividindo-o por n para encontrar a proporção de alunos do sexo masculino nesta amostra de pesquisa.

```
library(MASS)
sex = na.omit(survey$Sex)
n = length(sex)
k = sum(sex == "Male")
k
```

118

```
pbar  = k/n;pbar
```

0.5

O número de alunos do sexo masculino é 118. A proporção de alunos do sexo masculino é 0,5.

Agora, queremos encontrar o tamanho da amostra para atingir 5% de margem de erro para a pesquisa de estudantes do sexo masculino com nível de confiança de 95%

```
zstar = qnorm(0.975)
p=0.5

# Margem ou erro
E = 0.05
zstar^2*p*(1-p)/E^2
```

384.145882069412

Assim, concluímos que precisamos de 384,1459 ou 384 tamanhos de amostra para obter uma margem de erro de 5% para a pesquisa de estudantes do sexo masculino com um nível de confiança de 95%.