

Exercício 3 - Metodologia de Pesquisa em Computação

Gabriel Luciano Gomes (265673) Gabriel Oliveira Campos (265146)
Paulo Junio Reis Rodrigues (265674)

October 28, 2020

1 Amostras não-pareadas

```
# ----- EX3 - Part 1 ----- #
database <- read.csv("D:/dbs/MO430/ex1.csv")

database$type <- sapply(database$type, as.character)
database[which(database$type == 'No'), "type"] <- 0
database[which(database$type == 'Yes'), "type"] <- 1
database$type <- sapply(database$type, as.numeric)

bpNo <- array()
bpYes <- array()
y <- 1
z <- 1

for(i in 1:length(database$type)){
  if(database$type[i] == 0){
    bpNo[y] <- database$bp[i]
    y <- y+1
  } else {
    bpYes[z] <- database$bp[i]
    z <- z+1
  }
}

# ----- AMOSTRAS NAO PAREADAS ----- #
library(BEST)
bayesianCalcs = BESTmcmc(bpYes, bpNo)
summary(bayesianCalcs)
summary(bayesianCalcs, ROPE=c(-4,4))
```

```
> library(BEST)
> bayesianCalcs = BESTmcmc(bpYes,bpNo)
waiting for parallel processing to complete...done.
> summary(bayesianCalcs)
```

	mean	median	mode	HDI%	HDIlo	HDIup	compval	%>compval
mu1	74.587	74.588	74.578	95	71.862	77.407		
mu2	69.459	69.462	69.492	95	67.541	71.328		
muDiff	5.128	5.125	5.088	95	1.700	8.417	0	99.8
sigma1	11.061	11.007	10.880	95	8.887	13.356		
sigma2	10.528	10.514	10.453	95	8.939	12.093		
sigmaDiff	0.533	0.497	0.539	95	-1.947	3.084	0	65.5
nu	28.470	20.818	11.554	95	3.814	76.034		
log10nu	1.337	1.318	1.292	95	0.764	1.942		
effsz	0.477	0.476	0.477	95	0.158	0.795	0	99.8

Figure 1: Resultado dos testes Bayesianos - Amostras não pareadas

```
> summary(bayesianCalcs, ROPEm=c(-4,4))
```

	mean	median	mode	HDI%	HDIlo	HDIup	compval	%>compval	ROPElow	ROPEhigh	%InROPE
mu1	74.587	74.588	74.578	95	71.862	77.407					
mu2	69.459	69.462	69.492	95	67.541	71.328					
muDiff	5.128	5.125	5.088	95	1.700	8.417	0	99.8	-4	4	25.3
sigma1	11.061	11.007	10.880	95	8.887	13.356					
sigma2	10.528	10.514	10.453	95	8.939	12.093					
sigmaDiff	0.533	0.497	0.539	95	-1.947	3.084	0	65.5			
nu	28.470	20.818	11.554	95	3.814	76.034					
log10nu	1.337	1.318	1.292	95	0.764	1.942					
effsz	0.477	0.476	0.477	95	0.158	0.795	0	99.8			

Figure 2: Resultado dos testes Bayesianos com ROPE [-4,4] - Amostras não pareadas

1.1 Qual a probabilidade que a media do bp dos pacientes com diabetes é maior que a media dos pacientes sem?

Para analisar a probabilidade entre as amostras, μ_1 (bp dos pacientes com diabetes) e μ_2 (bp dos pacientes sem diabetes), é necessário efetuar o cálculo da média de cada uma delas. Para isto, foi utilizado o pacote *Bayesian Estimation Supersedes the t-Test* (BEST), que realiza estas operações necessárias. Além disso, essa biblioteca, junto à função *summary*, fornece um parâmetro "*compVal*", que é um indicador comparativo da diferença entre os parâmetros fornecidos (valor *default* = 0), que será utilizado para esta verificação. Por fim, deseja-se saber a probabilidade que a média de μ_1 seja maior que μ_2 e, para isto, basta verificar a porcentagem de elementos maiores que o $\% > compVal$. Para este exercício, obteve-se o valor de **99.8**, ou seja, a probabilidade dos casos de μ_1 serem maiores que μ_2 é de **99.8%**.

1.2 Assumindo um ROPE de -4 a 4, ou seja uma diferença na média menor que 4 é considerada como irrelevante. Qual a probabilidade que não há diferença prática ente a pressão sanguínea dos pacientes com e sem diabetes?

Como a diferença entre de -4, 4 na média é irrelevante, este parâmetro deve ser informado no cálculo do *compVal*. Para isto, utiliza-se o parâmetro *ROPEm* parâmetro da função *summary*. Após efetuar esta operação, ele retornará os valores *ROPElow*, *ROPEhigh* e *%InROPE*, que diz respeito ao limite inferior e superior da diferença e o percentual de dados estarem contidos dentro

de uma região (-4 e 4 da média). Feito isso, temos o resultado de **25.3**, ou seja, a probabilidade de que não há diferença prática entre as amostras é de **25.3%**.

2 Amostras pareadas

```
# ----- AMOSTRAS PAREADAS ----- #

database <- read.csv("D:/dbs/MO430/ex1-paired.csv")

Aug <- database$August
Nov <- database$November

library(BEST)
bayesianCalcs = BESTmcmc(Nov-Aug)
summary(bayesianCalcs)
summary(bayesianCalcs, ROPE=c(-3.5,3.5))
```

```
> library(BEST)
> bayesianCalcs = BESTmcmc(Nov-Aug)
waiting for parallel processing to complete...done.
> summary(bayesianCalcs)
```

	mean	median	mode	HDI%	HDIlo	HDIup	compVal	%>compVal
mu	4.518	4.474	4.43	95	-0.1610	9.25	0	97.3
sigma	7.896	7.603	7.16	95	4.1875	12.33		
nu	30.005	21.019	5.97	95	1.0099	87.75		
log10nu	1.283	1.323	1.46	95	0.4064	2.07		
effsz	0.604	0.599	0.61	95	-0.0407	1.22	0	97.3

Figure 3: Resultado dos testes Bayesianos - Amostras Pareadas

```
> summary(bayesianCalcs, ROPE=c(-3.5,3.5))
```

	mean	median	mode	HDI%	HDIlo	HDIup	compVal	%>compVal	ROPElow	ROPEhigh	%InROPE
mu	4.518	4.474	4.43	95	-0.1610	9.25	0	97.3	-3.5	3.5	32.8
sigma	7.896	7.603	7.16	95	4.1875	12.33					
nu	30.005	21.019	5.97	95	1.0099	87.75					
log10nu	1.283	1.323	1.46	95	0.4064	2.07					
effsz	0.604	0.599	0.61	95	-0.0407	1.22	0	97.3			

Figure 4: Resultado dos testes Bayesianos com ROPE [-3.5,3.5] - Amostras pareadas

2.1 Qual a probabilidade que a média do mês de Novembro seja maior que a media do mês de Agosto?

O pacote *BEST* não apresenta a possibilidade do cálculo para amostras pareadas. Entretanto, pode-se utilizar uma única amostra resultante da diferença entre as duas prévias. Com isso, pode-se calcular o *compVal* e obter o $\% > compVal$, semelhante ao apresentado no exercício 1.1. Como conseguinte, temos que $\% > compVal = 97.3$, então, a probabilidade de que a média do mês de Novembro seja maior do que a do mês de Agosto é de 97.3%.

2.2 Assumindo um ROPE de -3.5 a 3.5, qual a probabilidade da diferença não seja relevante?

Como apresentado no exercício 1.2, deve-se observar o valor de *%inROPE* obtivo pelo *summary* dentro do intervalo $[-3.5, 3.5]$ em relação à média. Sendo assim, obteve-se o valor de 32.8, ou seja, a probabilidade da diferença não ser relevante entre as amostras é de **32.8%**.