

# ME111 - Laboratório de Estatística

## Aula 5 - Aleatorização

Profa. Larissa Avila Matos

- Versão 1: Amedrontado por forças inimigas, um general enfrenta um dilema. Seus oficiais de inteligência dizem que seus soldados serão capturados em uma emboscada na qual 600 deles morrerão a não ser que ele os proteja escolhendo uma de duas rotas disponíveis. Se ele escolhe a primeira rota, 200 soldados serão salvos. Se ele escolhe a segunda, existe  $1/3$  de chance que 600 soldados serão salvos e  $2/3$  de chance que nenhum será salvo. Qual rota ele deveria tomar? (salvando vidas)

- Versão 2: Amedrontado por forças inimigas, um general enfrenta um dilema. Seus oficiais de inteligência dizem que seus soldados serão capturados em uma emboscada na qual 600 deles morrerão a não ser que ele os proteja escolhendo uma de duas rotas disponíveis. Se ele escolhe a primeira rota, 400 soldados morrerão. Se ele escolhe a segunda, existe  $1/3$  de chance que nenhum soldado morrerá e  $2/3$  de chance que os 600 soldados morrerão. Qual rota ele deveria tomar? (prevenindo mortes)

- Objetivo: o tratamento (forma de perguntar) influencia na resposta?
- Neste experimento completamente aleatorizado temos:
  - tratamentos (forma de perguntar: versão 1 e versão 2 são associados aos sujeitos de forma aleatória
  - $n = 14$  sujeitos para cada tratamento
  - Resposta: rota 1 ou rota 2

- No dilema do General, fatores além da versão lida (tratamento) podem influenciar a resposta dos sujeitos. Ex: sexo, religião, experiência militar, educação, renda, etc.
- Suponha que a versão 2 (tratamento 2) esteja relacionada com sujeitos que gostam de se arriscar.
- Suponha que a resposta seja completamente determinada pela personalidade dos sujeitos: todas as pessoas com personalidade avessa ao risco selecionam a rota 1 e aquelas que se arriscam selecionam a rota 2.
- O tipo da personalidade do sujeito pode ser determinado corretamente.

- Possíveis problemas podem acontecer:
  - Pesquisador não ético associa sujeitos avessos ao risco à primeira versão e todos os que se arriscam à segunda versão.
  - Dado que a versão lida não tem (verdadeiro) efeito na resposta os dados mostrariam que todos os sujeitos sob tratamento rota 1 selecionariam a rota 1 e aqueles sob tratamento rota 2 escolheriam a rota 2.
  - Suponha que os 14 indivíduos que chegam ao local do experimento recebam o tratamento 1. Se pessoas que chegam cedo são aquelas que não se arriscam, teríamos um forte vício!
  - Uma aleatorização infeliz pode acontecer, mas pode ser quantificada precisamente.

- Como atribuir tratamentos aos sujeitos de forma aleatória?
- Maneira mecânica:
  - Associamos os números 1 a 28 aos sujeitos de maneira arbitrária.
  - Pesquisador tem cartas (diferentes só por causa do número que ela contém) enumeradas de 1 a 28 que são depositadas em uma urna.
  - Sem olhar o pesquisador pega 14 cartas da urna.
  - Os sujeitos associados às cartas escolhidas recebem o tratamento correspondente a versão 1; os restantes à versão 2.
- Como aleatorizamos no R?

- Função `sample()` para Dilema do General ( $n_1 = n_2 = 14$ ).

```
x <- 1:28
```

```
x
```

```
[1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23  
[24] 24 25 26 27 28
```

```
sample.trt1 <- sample(1:28,14) # tratamento rota 1  
sample.trt1
```

```
[1] 24 15 14 21 5 17 1 6 7 27 3 2 16 25
```

```
sample.trt2 <- x[-sample.trt1] # tratamento rota 2  
sample.trt2
```

```
[1] 4 8 9 10 11 12 13 18 19 20 22 23 26 28
```



- Função `rarPar()` para Dilema do General ( $n_1 = n_2 = 14$ ).

```
require(randomizeR)
```

```
N <- 28  
params1 <- rarPar(N,groups=c("Versão1","Versão2"))  
R1 <- genSeq(params1)  
R1
```

Object of class "rRarSeq"

```
design = RAR  
seed = 159936  
N = 28  
groups = Versão1 Versão2
```

The sequence M:

```
1 Versão1 Versão1 Versão1 Versão1 Versão1 Versão1 Versão2 Versão2 Versão2 Versão2
```

```
seq1<-getRandList(R1)
seq1
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]
[1,]	"Versão1"	"Versão1"	"Versão1"	"Versão1"	"Versão1"	"Versão1"	"Versão2"
	[,8]	[,9]	[,10]	[,11]	[,12]	[,13]	[,14]
[1,]	"Versão2"	"Versão2"	"Versão1"	"Versão1"	"Versão2"	"Versão1"	"Versão2"
	[,15]	[,16]	[,17]	[,18]	[,19]	[,20]	[,21]
[1,]	"Versão2"	"Versão2"	"Versão1"	"Versão1"	"Versão2"	"Versão2"	"Versão2"
	[,22]	[,23]	[,24]	[,25]	[,26]	[,27]	[,28]
[1,]	"Versão2"	"Versão2"	"Versão1"	"Versão1"	"Versão2"	"Versão2"	"Versão1"

```
table(seq1)
```

```
seq1
Versão1 Versão2
      14      14
```

■ Função `crPar()` para Dilema do General (não balanceado)

```
N <- 28  
params2 <- crPar(N,groups=c("Versão1","Versão2"))  
R2 <- genSeq(params2)  
R2
```

Object of class "rCrSeq"

```
design = CR  
seed = 1004333424  
N = 28  
groups = Versão1 Versão2
```

The sequence M:

```
1 Versão2 Versão2 Versão1 Versão1 Versão2 Versão1 Versão2 Versão1 Versão2
```

```
seq2<-getRandList(R2)
seq2
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]
[1,]	"Versão2"	"Versão2"	"Versão1"	"Versão1"	"Versão2"	"Versão1"	"Versão2"
	[,8]	[,9]	[,10]	[,11]	[,12]	[,13]	[,14]
[1,]	"Versão1"	"Versão1"	"Versão1"	"Versão2"	"Versão1"	"Versão1"	"Versão1"
	[,15]	[,16]	[,17]	[,18]	[,19]	[,20]	[,21]
[1,]	"Versão2"	"Versão2"	"Versão1"	"Versão1"	"Versão2"	"Versão1"	"Versão2"
	[,22]	[,23]	[,24]	[,25]	[,26]	[,27]	[,28]
[1,]	"Versão1"	"Versão2"	"Versão1"	"Versão2"	"Versão2"	"Versão1"	"Versão2"

```
table(seq2)
```

```
seq2
Versão1 Versão2
      15      13
```

- Função `sample()` para Dilema do General ( $n_1 = 10$ ;  $n_2 = 18$ ).

```
x <- 1:28  
x
```

```
[1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23  
[24] 24 25 26 27 28
```

```
sample.trt1 <- sample(1:28,10) # tratamento rota 1  
sample.trt1
```

```
[1] 1 11 9 14 27 10 23 2 6 18
```

```
sample.trt2 <- x[-sample.trt1] # tratamento rota 2  
sample.trt2
```

```
[1] 3 4 5 7 8 12 13 15 16 17 19 20 21 22 24 25 26 28
```

- Suponha que desejamos realizar um estudo com 32 indivíduos e que temos dois tratamentos A e B. Como podemos aleatorizar esses indivíduos para metade deles receba o tratamento A e a outra metade o tratamento B?

## Gerando amostras

```
table(sample(c(1,2),10,prob=c(0.5,0.5),replace=T))/10
```

```
1    2  
0.5 0.5
```

```
table(sample(c(1,2),100,prob=c(0.5,0.5),replace=T))/100
```

```
1    2  
0.45 0.55
```

```
table(sample(c(1,2),1000,prob=c(0.5,0.5),replace=T))/1000
```

```
1    2  
0.49 0.51
```

```
table(sample(c(1,2),10000,prob=c(0.5,0.5),replace=T))/10000
```

1	2
0.506	0.494

```
table(sample(c(1,2),100000,prob=c(0.5,0.5),replace=T))/100000
```

1	2
0.50196	0.49804

```
table(sample(c(1,2),1000000,prob=c(0.5,0.5),replace=T))/1000000
```

1	2
0.499319	0.500681



## Jogando Dados

```
n<-10000  
s<-sample(c(1,2,3,4,5,6), n, replace = TRUE)  
table(s)/n
```

```
s  
      1      2      3      4      5      6  
0.1664 0.1640 0.1672 0.1619 0.1723 0.1682
```

- Suponha que tenhamos uma urna com 6 bolas vermelhas e 7 bolas azuis.
  - 1 Como podemos simular  $n$  retiradas dessa urna com reposição? Considere  $n = 10, 100, 1000, 10000$ .
  - 2 Como podemos simular retiradas dessa urna sem reposição?

