

ME111 - Laboratório de Estatística

Aula 6 - Estudo de simulação (Dilema do General)

Profa. Larissa Avila Matos

Dilema do General

- Objetivo geral da aula: Vamos ver como reproduzir o dilema do general várias vezes.

Relembrando:

- Análise de tratamentos: Experimento balanceado.
- Duas versões:
 - salvando vida (versão 1); ou
 - prevenindo morte (versão 2).
- Em cada uma das versões o general tem que escolher entre duas rotas possíveis (rota 1 ou rota 2).
- Cada versão do dilema é contado de forma diferente e queremos testar se o modo de contar influencia na escolha da rota.

- Usando os resultados apresentados no livro do ‘Wardrop (pág 14), temos que

	Rota 1 (Sucesso)	Rota 2 (Fracasso)	Total
Tratamento 1	7	7	14
Tratamento 2	1	13	14
Total	8	20	28

- Tratamento 1 = Versão 1; Tratamento 2 = Versão 2

- Portanto, nosso interesse é testar:

$$H_0 : T_1 = T_2 \quad \text{versus} \quad H_1 : T_1 > T_2,$$

ou equivalentemente,

$$H_0 : T_1 - T_2 = 0 \quad \text{versus} \quad H_1 : T_1 - T_2 > 0,$$

onde T_i é o número de indivíduos no tratamento i que seguiram a rota 1.

- Para isso precisamos encontrar a distribuição dessa diferença $(T_1 - T_2)$ e através de um teste de hipótese rejeitar ou não H_0 .
- Porém, não sabemos a distribuição da diferença.
- Através da simulação conseguimos chegar muito próximo da distribuição verdadeira.

- Em computação, simulação consiste em empregar formalizações em computadores, tais como expressões matemáticas ou especificações mais ou menos formalizadas, com o propósito de imitar um processo ou operação do mundo real. Desta forma, para ser realizada uma simulação, é necessário construir um modelo computacional que corresponda à situação real que se deseja simular. Fonte: Wikipédia
- Em determinados casos devemos repetir esse estudo várias vezes.

Estrutura de repetição: **for**

- As estruturas de repetição, executam a repetição de um conjunto de instruções enquanto uma determinada condição é verdadeira.
- No R, podemos usar algumas estruturas de repetição, usaremos o **for**.
- O comando 'for executa um conjunto de instruções, um determinado número de vezes de acordo com uma condição enquanto uma variável de controle é incrementada ou decrementada a cada passagem pelo “laço” (loop em inglês).

Exemplos

- Imprimindo os números de 1 a 10:

```
for(i in 1:10){  
  print(i)  
}
```

Exemplos

- Imprimindo os números de 1 a 10:

```
for(i in 1:10){  
    print(i)  
}
```

[1] 1

[1] 2

[1] 3

[1] 4

[1] 5

[1] 6

[1] 7

[1] 8

[1] 9

[1] 10

- Imprimindo 10 vezes a palavra oi:

```
for(i in 1:10){  
    print("oi")  
}
```

- Imprimindo 10 vezes a palavra oi:

```
for(i in 1:10){  
  print("oi")  
}
```

[1] "oi"

[1] "oi"

[1] "oi"

[1] "oi"

[1] "oi"

[1] "oi"

[1] "oi"

[1] "oi"

[1] "oi"

[1] "oi"

- Imprimindo os elementos de um vetor:

```
v<-20:25  
for(i in 1:length(v)){  
  print(v[i])  
}
```

[1] 20

[1] 21

[1] 22

[1] 23

[1] 24

[1] 25

- Somando todos os elementos de um vetor. Seja $v=c(1,10,15,20)$.

- Somando todos os elementos de um vetor. Seja $v=c(1,10,15,20)$.

```
v<-c(1,10,15,20)
x<-0
for(i in 1:length(v)){
  x<-x+v[i]
}
x
```

[1] 46

- Exercício: Como eu poderia reescrever esse código eliminando um passo, ou seja, i começando de 2.

- Somando todos os elementos de um vetor. Seja $v=c(1,10,15,20)$.

```
v<-c(1,10,15,20)
x<-0
for(i in 1:length(v)){
  x<-x+v[i]
}
x
```

[1] 46

- Exercício: Como eu poderia reescrever esse código eliminando um passo, ou seja, i começando de 2.

```
v<-c(1,10,15,20)
x<-v[1]
for(i in 2:length(v)){
  x<-x+v[i]
}
x
```

■ Exercício: Considere o vetor $v1=(2,4,6,8,10)$, como posso fazer para reescrevê-lo como $v2=(10,8,6,4,2)$?

■ Tenho: $v1=(2,4,6,8,10)$

■ Quero: $v2=(10,8,6,4,2)$

- Exercício: Considere o vetor $v1=(2,4,6,8,10)$, como posso fazer para reescrevê-lo como $v2=(10,8,6,4,2)$?

- Tenho: $v1=(2,4,6,8,10)$

- Quero: $v2=(10,8,6,4,2)$

```
v1<-c(2,4,6,8,10)
n<-length(v1)
v2<-rep(0,n)
for(i in 1:n){
  v2[i]<-v1[n+1-i]
}
v2
```

```
[1] 10  8  6  4  2
```


- Imprimindo os números de 1 a 10 (usando `while`):

```
count<-0
while(count<10){
  count<-count+1
  print(count)
}
```

[1] 1

[1] 2

[1] 3

[1] 4

[1] 5

[1] 6

[1] 7

[1] 8

[1] 9

[1] 10

- Contando quantos elementos no vetor é igual a um determinado valor.
- Seja $s=(1,0,2,3,4,5,1,2,3,2)$, queremos contar a quantidade de 1's que tem nesse vetor.

- Contando quantos elementos no vetor é igual a um determinado valor.
- Seja $s=(1,0,2,3,4,5,1,2,3,2)$, queremos contar a quantidade de 1's que tem nesse vetor.

```
s<-c(1,0,2,3,4,5,1,2,3,2)
x<-sum(s==1)
x
```

```
[1] 2
```

- Exercício: Como poderíamos fazer usando o `for`?

- Contando quantos elementos no vetor é igual a um determinado valor.
- Seja $s=(1,0,2,3,4,5,1,2,3,2)$, queremos contar a quantidade de 1's que tem nesse vetor.

```
s<-c(1,0,2,3,4,5,1,2,3,2)
x<-sum(s==1)
x
```

```
[1] 2
```

- Exercício: Como poderíamos fazer usando o `for`?

```
count<-0
for(i in 1:length(s)){
  if(s[i]==1) count<-count+1
}
count
```

```
[1] 2
```

Voltando ao Dilema do General

■ Simulação Dilema do General

```
# 1º passo
```

```
B = 10000
```

```
n1=n2=14
```

```
n=28
```

```
x=NULL
```

```
c(rep(1,8),rep(2,20))
```

```
[1] 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
```

```
s = sample(c(rep(1,8),rep(2,20)))
```

```
s
```

```
[1] 1 2 2 1 2 2 1 1 2 1 2 2 2 2 2 2 2 1 2 2 2 1 1 2 2 2 2
```

```

for (i in 1:B){
  # condicionando no numero de sucessos/fracassos observados
  s = sample(c(rep(1,8),rep(2,20)))
  x[i] = sum(s[1:n1]==1)
}
table(x)

```

```

x
  0    1    2    3    4    5    6    7    8
  7  159  870 2358 3193 2410  870  127   6

```

```

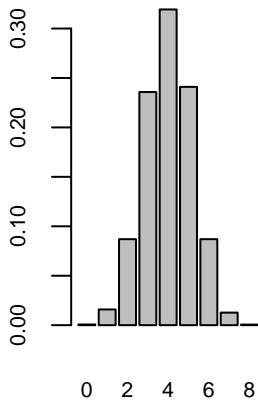
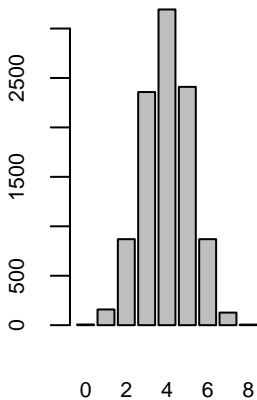
t<-table(x)/B
t

```

```

x
  0    1    2    3    4    5    6    7    8
0.0007 0.0159 0.0870 0.2358 0.3193 0.2410 0.0870 0.0127 0.0006

```



Distribuição exata

	R1	R2	
Tr1	0		14
Tr2	8		14
	8	20	28

$$T = \frac{0}{14} - \frac{8}{14} = -\frac{8}{14} = -\frac{4}{7}$$

	R1	R2	
Tr1	3		14
Tr2	5		14
	8	20	28

$$T = \frac{3}{14} - \frac{5}{14} = -\frac{2}{14} = -\frac{1}{7}$$

	R1	R2	
Tr1	6		14
Tr2	2		14
	8	20	28

$$T = \frac{6}{14} - \frac{2}{14} = \frac{4}{14} = \frac{2}{7}$$

	R1	R2	
Tr1	1		14
Tr2	7		14
	8	20	28

$$T = \frac{1}{14} - \frac{7}{14} = -\frac{6}{14} = -\frac{3}{7}$$

	R1	R2	
Tr1	4		14
Tr2	4		14
	8	20	28

$$T = \frac{4}{14} - \frac{4}{14} = \frac{0}{14} = 0$$

	R1	R2	
Tr1	7		14
Tr2	1		14
	8	20	28

$$T = \frac{7}{14} - \frac{1}{14} = \frac{6}{14} = \frac{3}{7}$$

	R1	R2	
Tr1	2		14
Tr2	6		14
	8	20	28

$$T = \frac{2}{14} - \frac{6}{14} = -\frac{4}{14} = -\frac{2}{7}$$

	R1	R2	
Tr1	5		14
Tr2	3		14
	8	20	28

$$T = \frac{5}{14} - \frac{3}{14} = \frac{2}{14} = \frac{1}{7}$$

	R1	R2	
Tr1	8		14
Tr2	0		14
	8	20	28

$$T = \frac{8}{14} - \frac{0}{14} = \frac{8}{14} = \frac{4}{7}$$

- Marginais fixas
- Conhecimento de uma casela determina as demais

	R1	R2	
Tr1	n_{11}		14
Tr2			14
	8	20	28

- Temos que, $Y_{11} \sim \text{Hipergeométrica}$, onde

$$\mathbb{P}(Y_{11} = n_{11}) = \frac{\binom{14}{n_{11}} \binom{14}{8 - n_{11}}}{\binom{28}{8}}, \quad n_{11} = 0, \dots, 8.$$

■ Ou seja, $T = \left\{ -\frac{4}{7}, -\frac{3}{7}, -\frac{2}{7}, -\frac{1}{7}, 0, \frac{1}{7}, \frac{2}{7}, \frac{3}{7}, \frac{4}{7} \right\}$

t	$\mathbb{P}(T = t)$	$\mathbb{P}(T \leq t)$	$\mathbb{P}(T \geq t)$
$-\frac{4}{7}$	0.001	0.001	1
$-\frac{3}{7}$	0.015	0.016	0.999
$-\frac{2}{7}$	0.088	0.104	0.984
$-\frac{1}{7}$	0.234	0.339	0.896
0	0.322	0.661	0.661
$\frac{1}{7}$	0.234	0.896	0.339
$\frac{2}{7}$	0.088	0.984	0.104
$\frac{3}{7}$	0.015	0.999	0.016
$\frac{4}{7}$	0.001	1	0.001
Total	1.000		

Distribuição exata no R

- Seja $m = n_1$ o número de sujeitos no tratamento 1, $n = n_2$ o número de sujeitos no tratamento 2 e k o número de sucessos total/rota1, então

```
dhyper(x=0,m=14,n=14,k=8)
```

```
[1] 0.0009661836
```

```
dhyper(x=1,m=14,n=14,k=8)
```

```
[1] 0.01545894
```

```
dhyper(x=2,m=14,n=14,k=8)
```

```
[1] 0.08792271
```

```
dhyper(x=3,m=14,n=14,k=8)
```

```
[1] 0.2344605
```

```
dhypcr(x=4,m=14,n=14,k=8)
```

```
[1] 0.3223833
```

```
dhypcr(x=5,m=14,n=14,k=8)
```

```
[1] 0.2344605
```

```
dhypcr(x=6,m=14,n=14,k=8)
```

```
[1] 0.08792271
```

```
dhypcr(x=7,m=14,n=14,k=8)
```

```
[1] 0.01545894
```

```
dhypcr(x=8,m=14,n=14,k=8)
```

```
[1] 0.0009661836
```

```
dhiper<-function(n){
  p1<-factorial(14)/(factorial(n) * factorial(14 - n))
  p2<-factorial(14)/(factorial(8-n) * factorial(6 + n))
  p3<-factorial(28)/(factorial(8) * factorial(20))
  return((p1*p2)/p3)
}

round(cbind(dhiper(0),dhiper(1),dhiper(2),dhiper(3),dhiper(4),
            dhiper(5),dhiper(6),dhiper(7),dhiper(8)),3)
```

```
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
[1,] 0.001 0.015 0.088 0.234 0.322 0.234 0.088 0.015 0.001
```

```

phiper<-function(x){
  p<-0
  if(x==0) p<-dhiper(x)
  else{
    p<-dhiper(0)
    for(i in 1:x){
      p<-p+dhiper(i)
    }
    return(p)
  }
round(cbind(phiper(0),phiper(1),phiper(2),phiper(3),phiper(4),
            phiper(5),phiper(6),phiper(7),phiper(8)),4)

```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]
[1,]	0.001	0.0164	0.1043	0.3388	0.6612	0.8957	0.9836	0.999	1

■ Simulação

```
# Distribuição de referência
x = NULL
B = 10000
n1=n2=14
n=28
for (i in 1:B){
  s = sample(c(rep(1,8),rep(2,20)))
  x[i] = sum(s[1:n1]==1)-sum(s[(n1+1):n]==1)
}
x.obs = 7 - 1
p.valor = mean(ifelse(x >= x.obs,1,0))
p.valor
```

```
[1] 0.016
```

■ Exato

```
p.valorE = (dhiper(7)+dhiper(8))  
p.valorE
```

```
[1] 0.01642512
```

■ Comparando

Simulação	Exato
0.016	0.0164

- Portanto, temos evidências ao nível de significância de 5% para rejeitar H_0 .

Teste Exato de Fisher

```
Dilema=matrix(c(7,1,7,13),nrow=2,  
              dimnames=list(Tratamento=c("Versão1","Versão2"),  
                             Rota=c("1-Sucesso1","2-Fracasso")))
```

Dilema

	Rota	
Tratamento	1-Sucesso1	2-Fracasso
Versão1	7	7
Versão2	1	13

```
# "greater" teste unilateral  
fisher.test(t(Dilema), alternative = "greater")
```

Fisher's Exact Test for Count Data

```
data:  t(Dilema)  
p-value = 0.01643  
alternative hypothesis: true odds ratio is greater than 1  
95 percent confidence interval:  
 1.493554      Inf  
sample estimates:  
odds ratio  
 11.82399
```

- Experimento ME110 - Aula Nancy
 - Faça um estudo de simulação reproduzindo o experimento que vocês realizaram na Aula de ME110. Considere $B = 10000$, calcule o p-valor.
 - Faça o teste Exato de Fisher

■ $H_0 : T_1 = T_2$ versus $H_1 : T_1 > T_2$,

```
x = NULL
B = 10000
n1=36
n2=35
n=71
for (i in 1:B){
  s = sample(c(rep(1,29),rep(2,42)))
  x[i] = sum(s[1:n1]==1)-sum(s[(n1+1):n]==1)
}
x.obs = 13 - 16
p.valor = mean(ifelse(x >= x.obs,1,0))
p.valor
```

```
[1] 0.8529
```

```
Dilema=matrix(c(13,16,23,19),nrow=2,  
              dimnames=list(Tratamento=c("Versão1","Versão2"),  
                             Rota=c("1-Sucesso1","2-Fracasso")))
```

Dilema

	Rota	
Tratamento	1-Sucesso1	2-Fracasso
Versão1	13	23
Versão2	16	19

```
fisher.test(t(Dilema), alternative = "greater")$p.value
```

```
[1] 0.8564472
```

Piada (Explicando o que é “looping”)

O diretor disse à secretária: – Vamos viajar para o exterior por uma semana, para um Seminário. Faça os preparativos da viagem!

A secretária faz uma chamada para o marido: – Vou viajar para o exterior com o diretor por uma semana. Cuide-se, querido.

O marido liga para a amante: – Minha mulher vai viajar para o exterior por uma semana, então nós vamos poder passar a semana juntos, meu docinho!

A amante liga para um menino a quem dá aulas particulares: – Tenho muito trabalho, na próxima semana não precisa vir às aulas.

O menino liga para o seu avô: – Vô, na próxima semana não tenho aulas, a minha professora estará ocupada. Vamos passar a semana juntos?!

O avô (que é o diretor desta história) liga para a secretária: – Vou passar a próxima semana com o meu neto, então não vou participar daquele Seminário. Pode cancelar a viagem.}

A secretária liga para o marido: – O diretor da empresa mudou de idéia e acabou cancelando a viagem.

O marido liga para a amante: – Não poderemos passar a próxima semana juntos, a viagem da minha mulher foi cancelada.

A amante liga para o menino das aulas particulares: – Mudança de planos: esta semana vamos ter aulas como normalmente.

O menino liga para o avô: – Vô, a minha professora disse que esta semana tenho aulas. Desculpe-me, não vai dar para fazer-lhe companhia.

O avô liga para a sua secretária: – Meu neto acabou de dizer que não vai poder ficar comigo essa semana. Continue com os preparativos da viagem ao seminário!

Entendeu o que é loop? Autor: Desconhecido.