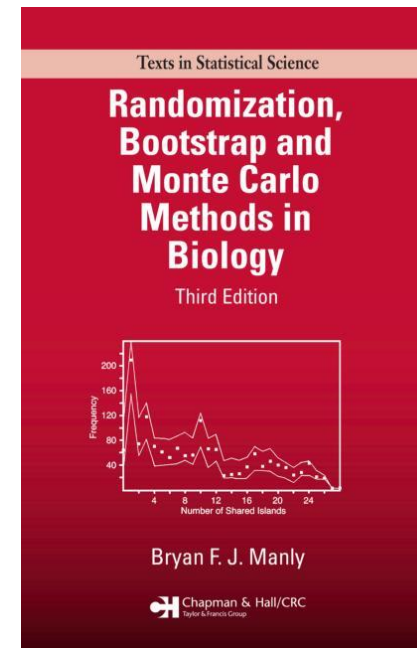


2

Aleatorização



- Manly, B. F. J. *Randomization, Bootstrap and Monte Carlo Methods in Biology*. 3ª ed., Chapman & Hall/CRC Press, Boca Raton, 2007.

ESQUEMA DO CAPÍTULO

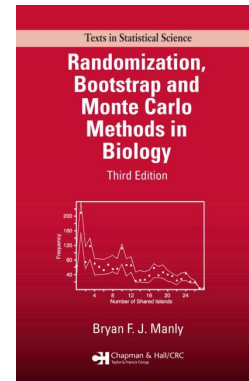
2.1 MÉTODOS DE ALEATORIZAÇÃO

2.2 MÉTODO JACKKNIFE

2.1 Métodos de Aleatorização

- Os **métodos de aleatorização** podem ser usados para:
 - Teste de hipóteses;
 - Intervalos de confiança (sob certas condições);
- Tais **métodos** envolvem:
 - a determinação do nível de significância de uma estatística de teste obtida do **conjunto de dados observado**,
 - por meio desta mesma estatística obtida de uma **permutação do conjunto de dados observado**;
- Estes são métodos **computacionalmente intensivos** uma vez que a aleatorização deve ser determinada ou
 - pela enumeração de **todas** as possíveis ordenações dos dados ou
 - pela obtenção de **uma amostra grande** dessas ordenações.

2.1 Métodos de Aleatorização (cont.)

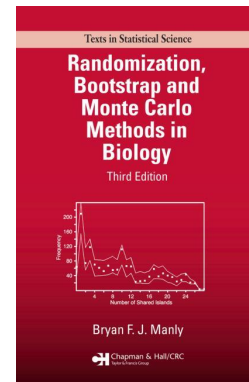


- Muitas hipóteses H_1 de interesse podem ser vistas como **alternativas** de hipóteses H_0 de aleatoridade;
- Ou seja, a hipótese sob investigação sugere que haverá uma tendência para um certo padrão nos dados, enquanto a hipótese nula diz que, se houver um padrão, então ele é efeito puramente do **acaso**;
- Testes de aleatorização são uma forma de determinar se a hipótese nula é **razoável** em tal situação;
- O valor s da estatística S para os dados observados é comparado com a distribuição de S que é obtida de uma **reordenação** aleatória dos dados;
- O argumento é que se a hipótese nula é verdadeira, então todas as possíveis ordenações dos dados são igualmente prováveis e s deverá aparecer como valor **típico** da distribuição de S .

2.1 Métodos de Aleatorização (cont.)

- O **valor P** de s é a proporção de valores que são tão extremos ou mais extremos que s na distribuição aleatória;
- Assim, se o **valor P** de s :
 - é menor que 5%, então há **alguma** evidência de que H_0 não é verdadeira;
 - é menor que 1%, então há uma evidência **forte** de que H_0 não é verdadeira;
 - É menos que 0,1%, então há uma evidência **muito forte** de que H_0 não é verdadeira.

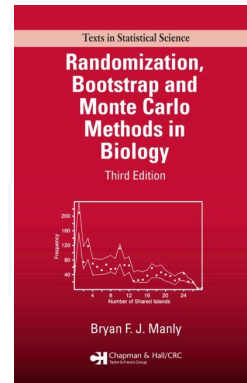
2.1 Métodos de Aleatorização (cont.)



Observações:

- De forma a **evitar** a caracterização da pessoa que faz análise como pertencente ao “*grupo de pessoas cujo objetivo na vida é de estar errado 5% do tempo*”,
É **melhor** considerar o nível de significância como:
 - uma medida da força da evidência contra H_0 ,
 - em vez de um medida que mostra se os dados são ou não significativos ao nível de 5%.

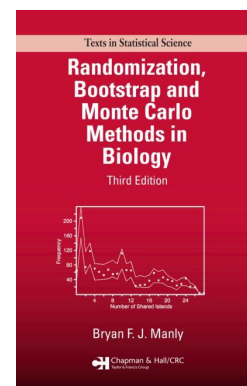
2.1 Métodos de Aleatorização (cont.)



Vantagem:

- Os métodos de aleatorização são válidos mesmo **sem** amostras aleatórias;

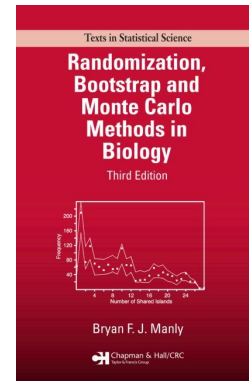
2.1 Métodos de Aleatorização (cont.)



Desvantagens:

- Nem sempre é possível **generalizar** as conclusões de um método de aleatorização para uma população de interesse;
- O que um método de aleatorização nos diz é se um certo padrão nos dados é ou não provável de ter ocorrido por **acaso**;
- O conceito de população da qual outras amostras poderiam ser retiradas é **desnecessário**, razão pela qual não é requerido pelos métodos de aleatorização que as amostras sejam aleatórias.

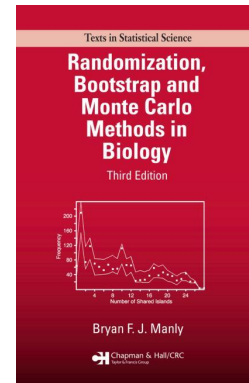
2.1 Métodos de Aleatorização (cont.)



Observações:

- Alguns estatísticos argumentam que
 - **falta teoria** para generalizar os resultados dos testes de aleatorização para uma população de interesse;
 - tais testes têm **pouco ou nenhum valor**, em comparação com testes clássicos para os quais existem métodos bem desenvolvidos de inferência estatística;
- Entretanto, outros argumentam que
 - a generalização dos resultados clássicos assume que as amostras são aleatórias, quando na realidade não são realmente aleatórias, mas simplesmente aquelas que estavam **disponíveis**.

2.1 Métodos de Aleatorização (cont.)



Exemplo 2.1.1:

Um fisiologista deseja verificar se o consumo moderado de álcool tem efeito no tempo de reação em jovens de 21 anos.

Em lugar de pegar uma amostra aleatória de todas as pessoas nessa idade, o que seria impossível, o/a fisiologista usa estudantes universitários nesta idade, matriculados na sua turma de Fisiologia.

Os estudantes são então divididos aleatoriamente em dois grupos:

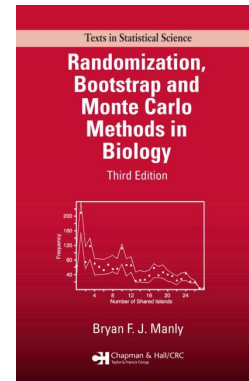
- um grupo tem os tempos de reação medidos após beberem uma pequena quantidade de **álcool** e
- outro grupo tem os tempos de reação medidos após beberem uma bebida **sem álcool**.

2.1 Métodos de Aleatorização (cont.)

Observações:

- **Vários** métodos poderiam ser utilizados para analisar os resultados desse experimento;
- Entretanto, conclusões baseadas nesse teste só serão válidas se assumirmos que os estudantes de Fisiologia são equivalentes a uma amostra **aleatória** de jovens de 21 anos, com respeito aos tempos de reação;
- Qualquer generalização nesse sentido requer o julgamento de que resultados similares são prováveis de ocorrer novamente, se um grupo **diferente** for testado;
- Claramente um experimento deste tipo **não dará** uma resposta definitiva, não importa quantas pessoas sejam utilizadas;

2.1 Métodos de Aleatorização (cont.)



Observações (cont.):

- Entretanto,
 - se o experimento for realizado em **outros** grupos (estudantes de direito, operários, secretários,...),
 - e os resultados forem sempre os **mesmos**,
 - então a maioria das pessoas acreditaria que os efeitos (ou falta de efeito) são **comuns** a todos os jovens de 21 anos.
- Este é o tipo de inferência que funciona **bem** com testes convencionais e testes de permutação.

2.1 Métodos de Aleatorização (cont.)

Exemplo 2.1.2:

Os dados abaixo mostram os comprimentos das mandíbulas (em mm) de chacais dourados (*Canis aureus*), para 10 de cada sexo, da coleção do Museu Britânico de História Natural, em Londres:

Machos:

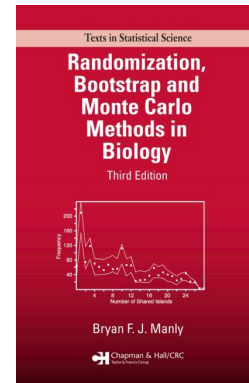
120 107 110 116 114 111 113 117 114 112

Fêmeas:

110 111 107 108 110 105 107 106 111 111



2.1 Métodos de Aleatorização (cont.)



Exemplo 2.1.2 (cont.):

- Os comprimentos foram medidos como parte de um estudo sobre a relação entre ossos caninos pré-históricos da Tailândia e ossos similares de espécies modernas;
- Para este exemplo, a questão de interesse é saber se há ou não evidência de diferenças entre os comprimentos médios para os dois sexos;
- Em outras palavras, há uma diferença entre as médias dos dois grupos de 4,8 mm;
- Entretanto, são realmente dois grupos ou este é um valor observado que poderia ser devido apenas ao acaso?

2.1 Métodos de Aleatorização (cont.)

Exemplo 2.1.2 (cont.):

- Dentre os métodos possíveis, será considerado na resolução do exemplo o teste- t com estimativa combinada (*pooled estimate*) do desvio padrão intragrupo;
- Suponha que os tamanhos das amostras sejam n_1 e n_2 , com médias e desvios-padrão $\bar{x}_1$, $\bar{x}_2$, s_1 e s_2 , respectivamente;
- Assuma que a amostra 1 seja proveniente de uma distribuição normal com média μ_1 e desvio-padrão σ , e
- a amostra 2 seja de uma distribuição normal, com média μ_2 e desvio-padrão σ ;

2.1 Métodos de Aleatorização (cont.)

Exemplo 2.1.2 (cont.):

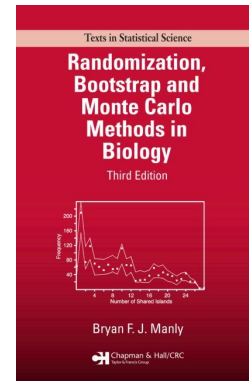
- A hipótese nula é $H_0: \mu_1 = \mu_2$, enquanto que a hipótese alternativa de interesse é $H_1: \mu_1 > \mu_2$;
- A hipótese nula é testada primeiramente obtendo-se a estimativa combinada da variância:

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

- É então calculada a estatística:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

2.1 Métodos de Aleatorização (cont.)



Exemplo 2.1.2 (cont.):

- Se a hipótese nula é verdadeira, então a estatística t será um valor aleatório da distribuição t com n_1+n_2-2 graus de liberdade;
- O teste t consiste em calcular a estatística t , para os dados do chacal dourado, e em verificar se o valor está dentro de uma faixa de valores que é razoável esperar apenas por acaso;
- Se o valor de t estiver nessa faixa, então a conclusão é que não há evidência contra a hipótese nula;
- Caso contrário, a hipótese nula deve ser abandonada e a hipótese alternativa pode ser considerada mais plausível.

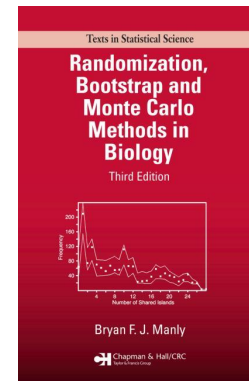
2.1 Métodos de Aleatorização (cont.)

Exemplo 2.1.2 (cont.):

- Realização do teste no R:

```
machos<-c(120,107,110,116,114,111,113,117,114,112)
femeas<-c(110,111,107,108,110,105,107,106,111,111)
t.test(machos, femeas, alternative=c("greater"), paired=FALSE,var.equal=TRUE)
```

2.1 Métodos de Aleatorização (cont.)



Exemplo 2.1.2 (cont.):

- Resultado:

Two Sample t-test

data: machos and femeas

$t = 3.4843$, $df = 18$, $p\text{-value} = 0.001324$

alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval:

2.411156 Inf

sample estimates:

mean of x mean of y

113.4 108.6

2.1 Métodos de Aleatorização (cont.)

Exemplo 2.1.2 (cont.):

- Para os **dados** fornecidos, tem-se que $x_{\text{barra1}}=113,4$ mm, $x_{\text{barra2}}=108,6$ mm, $s_1=3,72$ mm e $s_2=2,27$ mm;
- A estimativa **combinada** para o desvio padrão é $s = 3,08$ mm e $t = 3,4843$, com 18 graus de liberdade;
- Se a hipótese nula for verdadeira, a probabilidade de ocorrência de um valor assim tão **alto** é 0,001324;
- Portanto, o resultado é **significativo** ao nível de 0,1%, o que é geralmente visto como uma evidência forte contra a hipótese nula e, conseqüentemente, a favor da hipótese alternativa;

2.1 Métodos de Aleatorização (cont.)

Exemplo 2.1.2 (cont.):

As seguinte **suposições** foram feitas para esta análise:

1. Os dados são amostras **aleatórias** da população de interesse;
2. O desvio-padrão é o **mesmo** para os machos e as fêmeas;
3. Dentro de cada um dos grupos, os comprimentos das mandíbulas seguem uma distribuição **normal**.

2.1 Métodos de Aleatorização (cont.)

Exemplo 2.1.2 (cont.):

- A **suposição 1** é certamente **questionável**, pois os dados são provenientes de coleção de um museu, coletados de uma forma desconhecida;
- A **suposição 2** pode ser **verdadeira**; pelo menos os desvios-padrão dos grupos não diferem significativamente, em um teste F , pois

$$F = s_1^2/s_2^2 = 2,68,$$

com 9 e 9 graus de liberdade, não é significativamente diferente de 1,0 em comparação com o valor crítico de uma tabela F , com um nível de significância de 5%;

- A **suposição 3** pode ser **aproximadamente** verdadeira, mas não pode ser seriamente testada com amostras tão pequenas.

2.1 Métodos de Aleatorização (cont.)

Exemplo 2.1.2 (cont.):

- Considere agora o seguinte **teste de permutação**:

algoritmo

(1) calcule a diferença D_1 entre as médias dos machos e das fêmeas

repetir

(2) realoque aleatoriamente 10 observações ao grupo dos machos e as 10 restantes ao das fêmeas

(3) calcule a diferença D entre as médias dos machos e fêmeas

até número de **replicações** ser alcançado

(4) Se D_1 parece um valor típico de D , então conclua que não há diferenças entre as médias dos machos e das fêmeas

fim algoritmo

2.1 Métodos de Aleatorização (cont.)

Exemplo 2.1.2 (cont.):

- No passo (4) há a **questão** de D_1 parecer ou não um valor típico da distribuição aleatorizada D :
 - Se D_1 estiver entre 95% dos valores centrais da distribuição aleatorizada, então pode-se dizer que o resultado do teste é **não significativo**;
 - Se D_1 estiver entre 5% dos valores da cauda da distribuição aleatorizada, então o resultado é **significativo** ao nível de 5%;
 - Similarmente para 1% e 0,1%, que representam níveis crescentes de **significância** e de evidência contra a hipótese nula.

2.1 Métodos de Aleatorização (cont.)

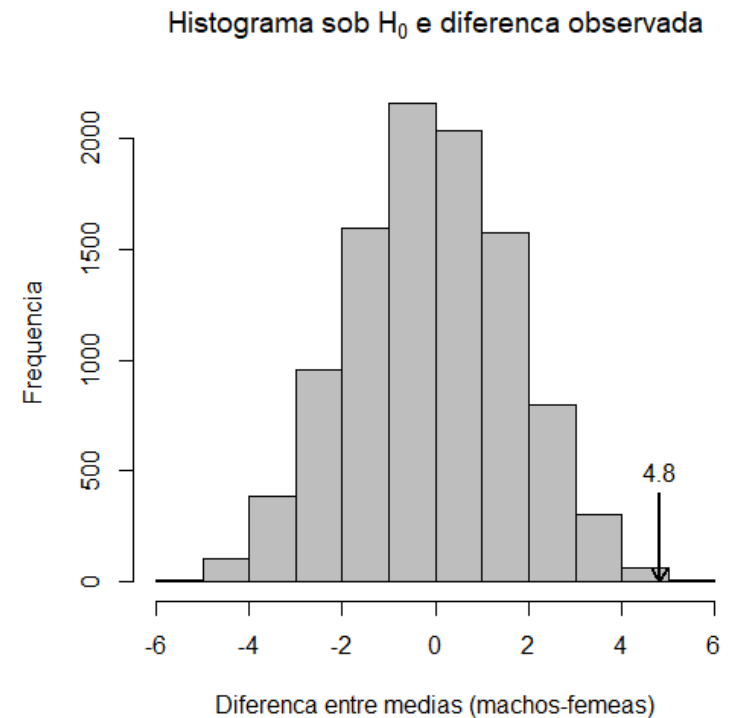
Exemplo 2.1.2 (cont.):

- Uma forma conveniente de sumarizar os resultados de um teste de permutação é calcular a **proporção** de valores observados para D que são maiores ou iguais a D_1 ;
- Tal proporção é uma estimativa da probabilidade de ocorrência de um valor tão extremo quanto D_1 , que é o **valor P** de D_1 .

2.1 Métodos de Aleatorização (cont.)

Exemplo 2.1.2 (cont.):

Histograma sob H_0 ,
gerado pelo **teste de
permutação**, para 10.000
replicações, e a diferença
entre médias de 4,8 mm,
com um **valor P =
0,0017**, como evidência
da **rejeição** de H_0 :



2.1 Métodos de Aleatorização (cont.)

Exemplo 2.1.2 (cont.):

- Como pode ser construído um **intervalo de confiança** de $(1-\alpha)100\%$ para a diferença entre médias, $L < \mu_D < U$, via teste de permutação?
- Os limites L e U podem ser encontrados por:
 - tentativa e erro ou
 - por uma busca sistemática;
- L é um valor que, subtraído da primeira amostra, produz um teste de permutação com **valor P** igual (ou próximo) de $\alpha/2$;
- U , subtraído da primeira amostra, produz no teste de permutação um **valor P** igual (ou próximo) de $1-\alpha/2$.

2.1 Métodos de Aleatorização (cont.)

Exemplo 2.1.2 (cont.):

- Suponha que você tenha disponível a função THPermut(machos,femeas), que receba os vetores machos e femeas e retorne o **valor P** do seguinte teste:
 - $H_0: \mu_{\text{machos}} - \mu_{\text{femeas}} = 0$ versus
 - $H_1: \mu_{\text{machos}} - \mu_{\text{femeas}} > 0$
- Iniciando-se com um $L = \mu_{\text{machos}} - \mu_{\text{femeas}}$, tente valores progressivamente mais baixos para L até que a chamada à função THPermut(machos,femeas) retorne um **valor P** em torno de $\alpha/2$ (por exemplo, $\alpha/2 = 0,025$, para um IC de 95% de confiança);
- Iniciando-se com um $U = \mu_{\text{machos}} - \mu_{\text{femeas}}$, tente valores progressivamente mais altos para U até que a chamada à função THPermut(machos,femeas) retorne um **valor P** em torno de $1-\alpha/2$ (por exemplo, $1-\alpha/2 = 0,975$, para um IC de 95% de confiança).

2.1 Métodos de Aleatorização (cont.)

Exemplo 2.1.2 (cont.):

- Para 10.000 replicações e um nível de 95% de confiança, tem-se a seguinte tabela para L:

L= 2.01	valor P= 0.0343
L= 2.00	valor P= 0.0343
L= 1.99	valor P= 0.0241
L= 1.98	valor P= 0.0241
- E, para U:

U= 7.65	valor P= 0.9739
U= 7.66	valor P= 0.9739
U= 7.67	valor P= 0.9774
U= 7.68	valor P= 0.9774
- O que sugere o seguinte limite de confiança de 95%, via teste de permutação:
[1,99; 7,66]
- Este limite é similar ao limite obtido via distribuição t , com 18 graus de liberdade, estimativa combinada do desvio padrão $s = 3,08$ e tamanhos de amostra n_1 e n_2 iguais a 10:
[1,91; 7,69]