

## Comparação de adesivos hormonais utilizando o método bootstrap pareado

## Comparison of hormonal adhesives using the paired bootstrap method

### RESUMO

Os métodos bootstrap, difundido por Efron e Tibshirani, são ferramentas que vem auxiliar os clássicos métodos de inferência estatística em questões ligadas a pressupostos exigidos. Neste estudo é realizado a implementação de um método bootstrap para amostras pareadas, o mesmo foi comparado ao método clássico t-pareado em relação a bioequivalência de dois adesivos hormonais. Será apresentado no decorrer do artigo gráficos e tabelas com análise descritivas dos dados referentes aos adesivos hormonais, bem como, métodos de execução e implementação dos testes t-pareado e t-bootstrap utilizando como auxílio, o software R open e pacotes desenvolvidos para estes fins. Como conclusão traremos os resultados obtidos com os métodos juntamente com nosso parecer em relação a utilização dos métodos t-pareado e t-bootstrap pareado na comparação dos adesivos hormonais.

**PALAVRAS-CHAVE:** Bioestatística. Amostragem. Estatística não-paramétrica.

### ABSTRACT

The bootstrap methods, disseminated by Efron and Tibshirani, are tools that help the classic methods of statistical inference in questions related to required assumptions. In this study, a bootstrap method for paired samples is implemented, which was compared to the classic t-paired method in relation to the bioequivalence of two hormonal adhesives. Graphs and tables with descriptive analysis of hormonal adhesive data will be presented during the article, as well as methods of execution and implementation of the t-paired and t-bootstrap tests using as an aid the R open software and packages developed for these purposes. In conclusion we will bring the results obtained with the methods together with our opinion regarding the use of the t-paired and paired t-bootstrap methods in the comparison of hormonal adhesives.

**KEYWORDS:** Biostatistics. Sampling. Non-parametric statistics.

**Gustavo de Medeiros**  
[Utfpr.medeiros@gmail.com](mailto:Utfpr.medeiros@gmail.com)  
Universidade Tecnológica Federal do Paraná, Toledo, Paraná, Brasil

**Gustavo Henrique Dalposso**  
[gustavodalposso@hotmail.com](mailto:gustavodalposso@hotmail.com)  
Universidade Tecnológica Federal do Paraná, Toledo, Paraná, Brasil

**Recebido:** 19 ago. 2020.

**Aprovado:** 01 out. 2020.

**Direito autoral:** Este trabalho está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.



## INTRODUÇÃO

Alvarado et al. (2017) nos falam sobre as ferramentas que a estatística nos proporciona, como os testes de hipótese e intervalo de confiança, utilizados amplamente em estudos biomédicos para avaliar a eficácia e bioequivalência de remédios, nos permitindo fazer inferências. Segundo Sheskin (2004) o teste mais utilizado nestes casos é o teste t de Student paramétrico, devido a ser robusto, porém o mesmo enfatiza os problemas que o teste pode trazer caso a amostra analisada seja pequena.

Efron e Tibshirani (1994) contudo, trazem um método de inferência estatística baseado em métodos computacionais e que, diferentemente do clássico teste t paramétrico não depende de pressupostos, como a normalidade dos dados ou tamanho da amostra. O Bootstrap, oriundo da frase sustentar-se pelos **bootstrap**<sup>1</sup>, das aventuras do barão Munchausen, foi pensado para criar inferências a partir de conjuntos de dados reais independente de pressupostos utilizando para isto reamostragens do conjunto de dados.

No Brasil a Agência Nacional de Vigilância Sanitária (ANVISA) que recebe informações diretamente da *Food and Drug Administration* (FDA), órgão fiscalizador da bioequivalência de remédios nos Estados unidos, trabalha frequentemente com métodos estatísticos a fim de testar a eficácia de produtos genérico em relação a seus concorrentes, visto que um remédio genérico só pode ser comercializado caso seus efeitos sejam os mesmos do remédio original.

Para isto, são realizados testes emparelhados com os dois medicamentos, onde o mesmo indivíduo, ou um casal selecionado, utiliza os medicamentos e fornecem amostras para estudos de bioequivalência. Os dados emparelhados ou combinados são definidos por Mann (2017) como um par de amostras A e B coletados da mesma fonte, em diferentes circunstâncias, onde para cada elemento da amostra A existe um único correspondente na amostra B.

Neste estudo estatístico, para fazer uma análise da eficácia dos métodos bootstrap, será determinado um intervalo de confiança sobre um conjunto de dados referente a bioequivalência de dois medicamentos, retirados do livro *An introduction to the Bootstrap* de Efron e Tibshirani (1994). O mesmo será comparado com um intervalo obtido através do clássico teste t-pareado, realizando discussões sobre a fidedignidade dos resultados obtidos.

O t-bootstrap pareado, por requerer poder computacional visto que os cálculos por mais simples que sejam, são repetitivos, nos fizeram recorrer ao *software* R. O script selecionado para se trabalhar com o R pode ser encontrado em Chihara (2019), que também carrega junto a si uma explicação sobre intervalos de confiança e testes de hipótese para amostras pareadas.

## MATERIAL E MÉTODOS

Os dados utilizados neste trabalho são referentes a uma pesquisa com adesivos hormonais, sendo um adesivo de uma marca antiga já disponível no mercado e aprovado pela FDA e outro de uma marca genérica que estava em

---

<sup>1</sup> O termo bootstrap refere-se a um laço de couro ou tecido costurado na parte traseira das botas para facilitar a colocação

processo de avaliação (EFRON & TIBSHIRANI, 1994, grifo nosso). Cada sujeito teve seus níveis sanguíneos do hormônio medidos após o uso dos adesivos (Tabela 1) e o interesse é investigar se os novos adesivos são igualmente eficazes aos adesivos já aprovados.

Tabela 1 – Níveis sanguíneos em indivíduos que utilizaram os adesivos hormonais.

Indivíduos	1	2	3	4	5	6	7	8
Adesivo Antigo	17649	12013	19979	21816	13850	9806	17208	29044
Adesivo Novo	16449	14614	17274	23798	12560	10157	16570	26325

Fonte: Efron & Tibshirani (1994).

Com objetivo de organizar, resumir e descrever os aspectos importantes do conjunto de dados da Tabela 1, serão calculadas algumas medidas de tendência central, variabilidade, assimetria e curtose. Para comparar graficamente com respeito à média, à dispersão e à distribuição os dados referentes ao adesivo antigo e o adesivo novo, será elaborado um gráfico boxplot, além de ser aplicado nos mesmos o teste t-pareado clássico e o t-bootstrap.

Para realizar o teste T *student* pareado, seguimos os seguintes passos:

**Passo 1:** primeiramente devemos indicar nossas hipóteses nula  $H_0$  e alternativa  $H_1$ , que dizem o valor do  $\mu_d$ , diferença das médias  $\mu_1$  e  $\mu_2$  em uma amostra de tamanho  $n$ .

$H_0: \mu_d = 0$  ( $\mu_1 - \mu_2 = 0$  não há diferença estatística entre as médias amostrais)

$H_1: \mu_d \neq 0$  ( $\mu_1 - \mu_2 \neq 0$  as médias das amostras diferem)

**Passo 2:** Verificar os pressupostos exigidos pelo método pois nem sempre ele pode ser aplicado, tendo um número amostral pequeno ( $n < 30$ ), não podemos trabalhar com a tabela de distribuição normal, porém, se for possível verificar que a população das diferenças é aproximadamente normal, segundo Mann (2017) faz sentido utilizar a distribuição t para fazer o teste.

**Passo 3:** Determinar a área de aceitação e rejeição do teste de hipótese, para isso será necessário uma tabela de distribuição t, trabalhando com  $(n - 1)$  graus de liberdade e um nível de significância ( $\alpha = 10\%$ ), conforme feito inicialmente por Efron e Tibshirani (1994), visto que é a norma estipulada pela FDA. Levando em consideração um teste bicaudal encontraremos os valores do t tabelado  $t_t$ , onde  $-t_t$  e  $t_t$  delimitam a área de aceitação.

**Passo 4:** Encontramos o t calculado  $t_c$ , através dos dados da Tabela 1.

A estatística amostral é dada por:

$$t_c = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}} \quad (1)$$

Onde  $d$  são as diferenças emparelhadas,  $\bar{d}$  é a média das diferenças emparelhadas e  $s_d$  é o desvio padrão, calculados respectivamente como:

$$\bar{d} = \frac{\sum d}{n} \quad (2)$$

$$S_d = \sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n-1}} \quad (3)$$

**Passo 5:** Queremos saber se as médias são iguais ou diferentes, a região delimitada pelos valores críticos  $t_t$  é a área de aceitação, nela aceitamos  $H_0$  a um nível  $\alpha$  de significância caso contrário, rejeitamos  $H_0$  e automaticamente aceitamos  $H_1$ . Comparam-se os valores do t tabelado com o resultado obtido para o t calculado, verificando se sua posição se encontra na área de aceitação ou rejeição da hipótese nula, caso esteja na área de aceitação podemos falar que os remédios possuem bioequivalência, ou são estatisticamente iguais.

Para realizarmos o teste com o bootstrap pareado, seguimos o modelo de algoritmo de Chihara (2019):

**Passo 1:** Selecionar as amostras pareadas e obter as suas diferenças.

**Passo 2:** Criar uma matriz para armazenar a média das réplicas obtidas a partir do conjunto das diferenças. Neste estudo realizamos o processo de reamostragem mil vezes, obtendo réplicas com o mesmo tamanho do conjunto amostral com repetições.

**Passo 3:** Criar mil reamostragens do conjunto das diferenças, calculando a média para cada um deles e armazena-las em uma matriz.

**Passo 4:** Encontrar o IC de Efron, ou intervalo percentil de Efron do conjunto das médias. Respeitamos as normas da FDA com um nível de significância ( $\alpha = 10\%$ ).

Descartamos os 5% das médias mais baixas e mais altas, o mínimo e o máximo das médias que restaram delimitam o intervalo percentil de Efron ou intervalo de confiança bootstrap.

Caso o 0 esteja contido neste intervalo, podemos assumir que há evidências estatísticas de que a média dos dois adesivos antigo e novo são iguais, fato equivalente a aceitar a hipótese nula em um teste t-pareado.

Se o 0 não fizer parte do intervalo percentil de Efron falamos que há evidências estatísticas de que há uma diferença entre as médias dos adesivos a um nível  $\alpha$  de significância, ou que os dois adesivos não possuem bioequivalência.

As análises foram desenvolvidas no *software* Microsoft R Open (Microsoft R Core Team, 2017). As medidas de tendência central e variabilidade foram implementadas pelos autores e as medidas de assimetria e curtose foram calculadas com auxílio das funções **skewness** e **kurtosis** do pacote e1071 (MEYER et al, 2019).

O gráfico boxplot foi elaborado pelos autores utilizando o comando **boxplot**, o teste t para dados pareados foi calculado com auxílio do da função **t.teste** do pacote stats (R Core Team, 2019). O histograma gerado a partir dos dados com o comando **hist** juntamente com o intervalo de confiança bootstrap com o comando **quantiles** bem como o método bootstrap pareado foram implementados pelos autores.

## RESULTADOS E DISCUSSÕES

Podemos observar que os dados obtidos com os adesivos novos estão de um modo geral mais próximos de sua média do que os dados obtidos com o adesivo antigo como podemos notar através do desvio padrão e do coeficiente de variação, além de que, ambos os adesivos apresentam uma assimetria positiva e muito semelhante (Tabela 2).

Analisando os dados é possível notar algumas relações entre a eficácia de ambos os adesivos, primeiramente o nível sanguíneo de indivíduos que utilizaram o adesivo antigo apresenta um mínimo e primeiro quartil menores do que o nível sanguíneo de indivíduos que utilizaram o adesivo novo, além de ter uma amplitude maior nos dados, ou seja, o efeito dos adesivos novos pode ser melhor do que os antigos, partindo do ponto de vista que a eficácia mínima é maior.

A média e mediana bem como o terceiro quartil de ambos por outro lado nos dizem o contrário, os adesivos antigos apresentam os maiores valores, a média elevada graças ao máximo dos adesivos antigos, sendo assim o efeito máximo é melhor nos adesivos antigos.

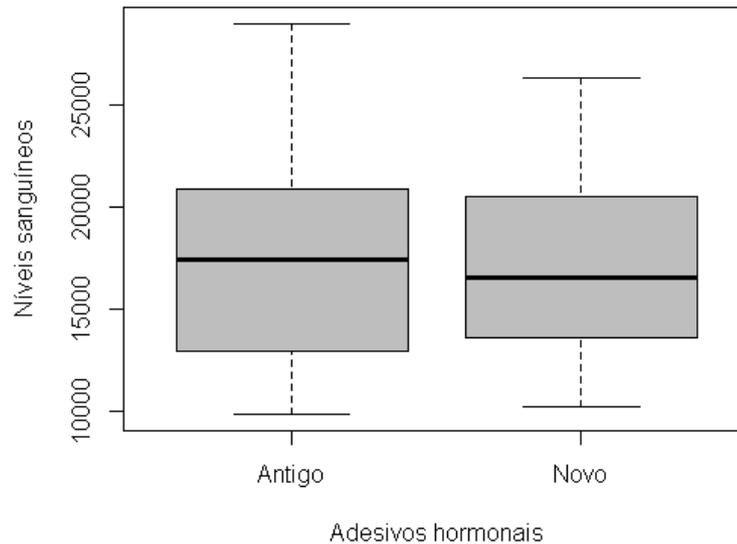
Tabela 2 – Estatísticas descritivas dos níveis sanguíneos em indivíduos que utilizaram os adesivos hormonais

Estatística	Adesivo Antigo	Adesivo Novos
Mínimo	9806	10157
Primeiro Quartil	13390,75	14100,5
Mediana	17428,50	16509,5
Média	17670,62	17218,38
Terceiro Quartil	20438,25	18905,0
Desvio Padrão	6100,809	5419,327
Assimetria	0,4576286	0,4464641
Curtose	-1,004119	-1,296406
Coeficiente de Variação	0,3452515	0,3147409
Máximo	29044	26325

Fonte: Dos autores (2020).

Observando os gráficos boxplots (Figura 1) fica evidente que a amplitude dos adesivos antigos é maior do que a dos adesivos novos, tendo assim um efeito mais volátil, podendo ser melhor, com um efeito máximo que os novos não atingirão, ou um pouco piores com um efeito levemente mais baixo.

Figura 1 – Gráficos boxplots dos níveis sanguíneos do hormônio.



Fonte: Dos autores (2020).

Porém, para que o adesivo novo seja aceito não basta ele apresentar um resultado mínimo maior, é necessário que o mesmo não apresente uma diferença estatística em relação ao efeito médio dos adesivos antigos. Com a tabela 2 vemos que a média do adesivo antigo é superior à do adesivo novo, mas não sabemos se isto significa de fato uma diferença a um nível de significância ( $\alpha = 10\%$ ).

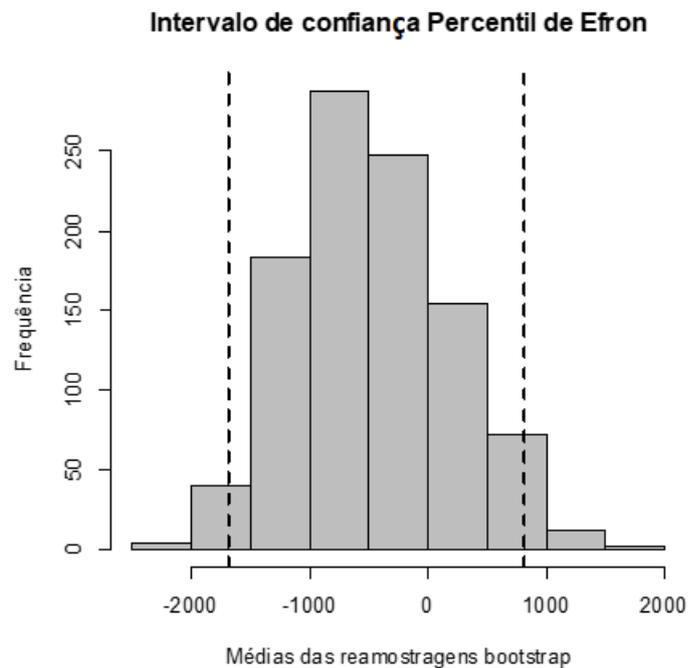
Com o teste t-pareado e a análise do teste de hipótese feito com ele, constatamos que os dois adesivos hormonais possuem bioequivalência. Para isto assumimos que todos os pressupostos para utilizar o teste t foram atendidos, obtendo os valores do t calculado e do t tabelado através do software R, como sendo respectivamente,  $-0,6465$  e  $1,8946$ .

Como  $-t_t < t_c < t_t$ , ou seja, o t calculado está entre os valores críticos, dentro da área de aceitação da hipótese nula, e assim não há diferença entre a média da eficácia dos remédios, porém como a quantidade de dados na amostra é pequena, sendo assim difícil verificar se a distribuição é de fato próxima a normal é prudente utilizar métodos bootstrap para obtermos uma comprovação.

Sabemos que a amostra original se aproxima da população a qual foi extraída, as reamostragens desta amostra feitas no método bootstrap aproxima o que teríamos se fosse possível retirar muitas amostras da população original, além de que a distribuição obtida se aproxima da normal conforme aumentamos o número de reamostragens.

Por fim o intervalo de confiança nos fornece o necessário para realizarmos a análise sobre a eficácia dos adesivos, demarcado na figura 2 através das linhas pontilhadas. Como o 0 está dentro do intervalo não podemos afirmar que há diferença na eficácia dos remédios, ou em outras palavras, eles possuem bioequivalência.

Figura 2 – Histograma da reamostragem das médias das diferenças e marcação do intervalo percentil de Efron



Fonte: Dos autores (2020).

## CONCLUSÕES

O intervalo de confiança fornecido pelo método bootstrap foi [-1959.759, 641.725] e a média foi aproximadamente -521, o fornecido pelo clássico método t pareado foi [-1777.5271, 873.0271] com uma média aproximada de -452. A amplitude dos dois intervalos foi muito próxima, e dos dois modos chegamos a mesma conclusão em relação aos adesivos hormonais, a bioequivalência existe, porém não podemos esquecer que assumimos que todas os pré-requisitos do teste clássicos foram aceitos, mesmo tendo uma amostra pequena, o que pode nos levar a resultados não tão fidedignos quanto os métodos de reamostragens bootstrap recomendados nestes caso, ou seja, os é mais aconselhável e prático utilizar o método bootstrap pareado.

## AGRADECIMENTOS

Os autores agradecem o apoio da Universidade Tecnológica Federal do Paraná, que tornou possível a realização deste trabalho.

## REFERÊNCIAS

ALVARADO, L. A.; DWIVEDI, A. K.; MALLAWAARACHCHIB, I. Analysis of small sample size studies using nonparametric bootstrap test with pooled resampling method, **Wiley Online Library**, 9 mar. 2017, ISSN: 2187–2205

CHIHARA, L. M. **Mathematical Statistics with Resampling and R**. Carleton College. Editora Wiley. 2º edição. 2019

EFRON, B.; TIBSHIRANI, R. J; **An introduction to the Bootstrap**. Editora CRC. 2º edição. 1994

MANN, P. S. **Introductory Statistics**. Connecticut state universit. Editora Wiley. 9º edição. 2017

MEYER, D.; DIMITRIADOU, E.; HORNIK, K.; WEINGESSEL, A. and LEISCH, F. (2019). **e1071**: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.7-1.

**R Core Team (2019)**. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

S original, from StatLib and by Rob Tibshirani. R port by Friedrich Leisch. (2017). **bootstrap**: Functions for the Book "An Introduction to the Bootstrap". R package version 2017.2.

SHEKIN, D. J. **Handbook of parametric and nonparametric statistical procedures**. Connecticut Ocidental universit. Editora CRC. 3º edição. 2004