

Bootstrap e Testes de Aleatorização

Métodos Estatísticos em Pesquisa Científica (MEPC)

Prof. Me. Lineu Alberto Cavazani de Freitas

Departamento de Estatística
Laboratório de Estatística e Geoinformação



Introdução

- ▶ O **avanço computacional** gerou um interesse e possibilidade de trabalhar com **inferência** estatística usando **métodos computacionalmente intensivos**.
- ▶ Estes métodos se baseiam em **reamostragem** e/ou **simulação**.
- ▶ São úteis quando as abordagens tradicionais não são adequadas por motivos como:
 - ▶ Pequenas amostras.
 - ▶ Violação de pressupostos.
 - ▶ Problemas “novos” para os quais não foi desenvolvido um procedimento.

Introdução

- ▶ **Reamostrar** significa **amostrar várias vezes** uma série de valores a partir dos dados observados.
- ▶ Dois procedimentos importantes são baseados em reamostragem: **Bootstrap** e o **Teste de Permutação**.
- ▶ O **Bootstrap** é uma interessante ferramenta que permite **construir uma distribuição amostral** e usar esta distribuição para fazer inferência.
- ▶ Os **Testes de Permutação** tem sua principal aplicação em problemas que envolvem **testes de hipóteses de comparação entre grupos** (diferença entre médias, proporções, variâncias, correlações, etc). Permite construir a distribuição sob H_0 .



Bootstrap

Bootstrap

- ▶ A ideia do Bootstrap é **simular o processo do paradigma frequentista** porém fazendo uso de uma única amostra.
- ▶ Ao invés de coletar diversas amostras, geramos reamostras da amostra original de mesmo tamanho, tomada com reposição.
- ▶ Para cada amostra, calcula-se a quantidade de interesse, gera-se a distribuição amostral e usa esta distribuição para fazer inferência (obter estimativas pontuais, intervalares, testar hipóteses).
- ▶ Não se baseia em teoremas ou outras suposições distribucionais.
- ▶ O número de repetições é arbitrário, mas quanto mais iterações forem feitas, mais precisas são as estimativas pontuais e intervalares.

Bootstrap

- ▶ A técnica ganhou popularidade com os artigos e livros de Bradley Efron.
- ▶ Quando proposto, enfrentou considerável ceticismo.
- ▶ O método não compensa pequenos tamanho de amostra, não cria novos dados nem preenche lacunas. Ele meramente parte do pressuposto que a amostra original é representativa, e gera novas amostras com variações.
- ▶ Na Estatística moderna, Bootstrap tem um papel importante para avaliar propriedades de estimadores.
- ▶ Na modelagem, bootstrap pode ser usado gerando conjuntos de dados como um todo, basta reamostrar as linhas.



Testes de permutação/aleatorizados

Testes de permutação/aleatorizados

- ▶ Podem ser usados para **testar hipóteses, geralmente envolvendo grupos.**
- ▶ O procedimento consiste em combinar duas ou mais amostras e aleatoriamente realocar as observações em reamostras.
- ▶ Pode ser feito de forma exaustiva, isto é, obter todas as possíveis combinações.
- ▶ Os testes de permutação exaustiva são práticos para tamanhos de amostra relativamente pequenos e são chamados de testes exatos. Para tamanhos grandes de amostra se tornam inviáveis

Testes de permutação/aleatorizados

De modo geral:

1. Gere um único grupo com as observações de todos os grupos.
2. Extraia uma amostra de tamanho igual ao tamanho do grupo A.
3. Dos dados restantes, faça o mesmo procedimento para o grupo B.
4. Se houver mais grupos, repita o procedimento.
5. Calcule a estatística de interesse baseada nos dados da reamostra (ex: diferença entre proporções do grupo A e B) e registre.
6. Repita o procedimento e construa uma distribuição de probabilidades empírica. Esta é a distribuição sob hipótese nula.
7. Veja onde o valor observado se encontra nesta distribuição. Se o valor observado se encontrar na área mais densa da distribuição, não há evidência de diferença, do contrário há.