BOOTSTRAP
Tomada de decisão via Intervalo de Confiança para alguma medida populacional

Magalhães e Lima - Seção 7.4 Concentrar-se em intervalo de confiança para média POPULACIONAL

Bootstrapping - https://www.thoughtco.com/example-of-bootstrapping-3126155

Bootstrapping - https://www.stat.auckland.ac.nz/~wild/BootAnim/

Objetivos de Aprendizagem

Os alunos devem ser capazes de:

- ✓ **Entender metodologia de Bootstrapping e usá-la para tomada de decisão via construção de Intervalo de Confiança para alguma medida populacional com um determinado coeficiente de confiança.**

Acompanhe, previamente, o PLANO DE AULA no BLACKBOARD!

Motivação:

Uma gerente de uma fábrica decidiu recalibrar uma das máquinas da linha de produção para AJUSTAR o seu tempo médio de execução. A amostra abaixo representa o tempo medido entre a entrada do bloco cru e a saída manufaturada em minutos.

In [2]:
```
x = [0.69689142,  0.57995802,  0.06450754,  0.61206626,  0.91976314,  0.34363861,  0.82575631,
     0.41666036,  0.62153527,  0.56757104,  0.86683518,  0.28922458,  0.28521698]

n = len(x)

print("Tamanho da Amostra: {0}".format(n))
print("Média: {0}".format(np.mean(x)))
print("Desvio Padrão: {0}".format(np.std(x, ddof=1)))
```

```
Tamanho da Amostra: 13
Média: 0.5453557469230769
Desvio Padrão: 0.2547555304406469
```

Problema:

Excepcionalmente hoje, os funcionários estavam reclamando que o tempo médio de execução da linha de produção está com comportamento diferente do usual, pois o tempo de execução histórico demora em média 0.68 minutos.

Avalie se a reclamação dos funcionários procede, com 95% de confiança?

Qual técnica inferencial utilizar para resolver tal problema?

Considerações do que temos até o momento:

Como visto em aulas passadas, para poder realizar um Teste de Hipóteses clássico quando o problema questiona a média populacional, ou seja, μ, temos as seguintes premissas (suposições):

1. **Primeiro resultado:**

    - A variável de interesse X tiver distribuição **EXATA** normal, então **cada** X_i **é normal** e a distribuição da média amostral $\overline{X} = \frac{X_1+X_2+...+X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$ é **EXATA** Normal.
    - A variável de interesse X tiver distribuição desconhecida ou não normal, então **cada** X_i **não é normal** e a distribuição da média amostral $\overline{X} = \frac{X_1+X_2+...+X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$ é **APROXIMADA** Normal, via TLC, **considerando** n **suficientemente grande**.

2. **Segundo resultado:**

    - Quando a **variância populacional** σ^2 **é conhecida**, pode-se usar:

      $\Rightarrow \overline{X} \sim N(\mu, \sigma^2/n) \Rightarrow \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$, quando X tem normal ou quando n é suficientemente grande.

    - Quando a **variância populacional** σ^2 **é DESconhecida**, pode-se usar:

      $\Rightarrow \frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$, quando X tem normal ou quando n é suficientemente grande, sendo S o desvio padrão AMOSTRAL.

Resumindo: Quando não sabemos qual é a distribuição de X e nem podemos assumir que n é suficientemente grande, uma nova técnica se faz necessária!!!

Bootstrapping

Solução: Métodos não paramétricos

Bootstrapping é um método de reamostragem com reposição que consiste em recuperar a distribuição de uma medida de interesse a partir de uma amostra mestre. Por exemplo, podemos utilizar a média como medida de interesse, mas a distribuição de outras medidas também pode ser estimadas, como por exemplo variância, mediana, máximo, etc.

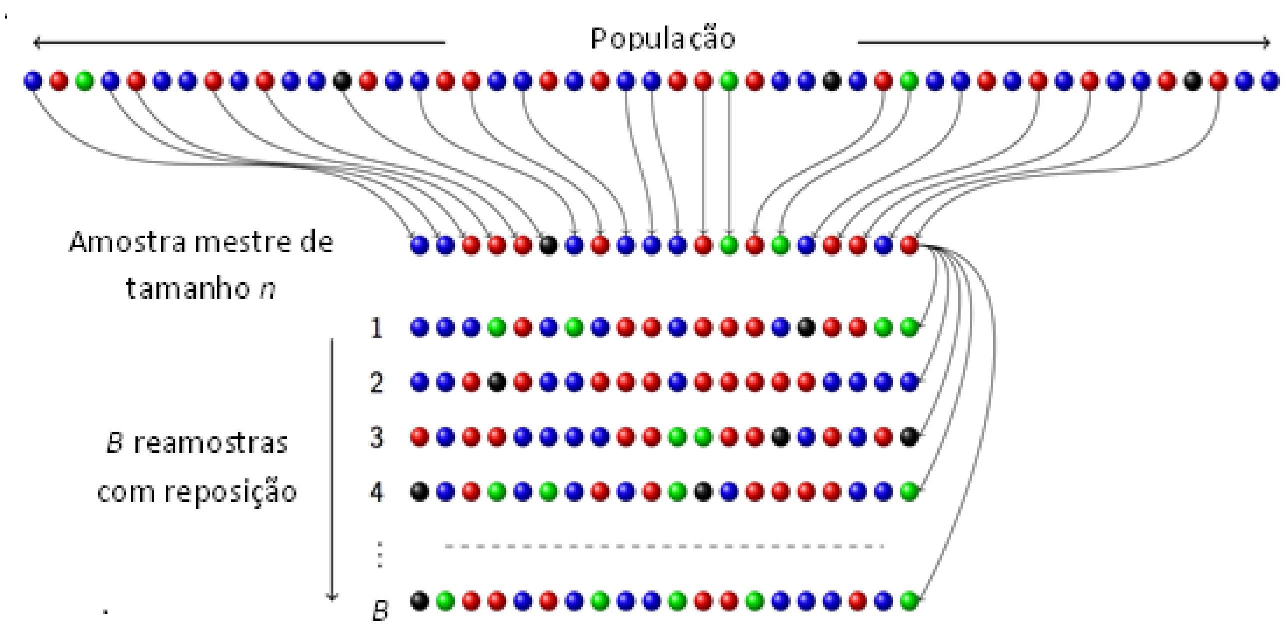
Reamostragem com reposição significa que um mesmo elemento pode ser selecionado várias vezes, assim como outro elemento pode não ser escolhido. Seria equivalente a sortear um número e colocar de volta na urna.

Como funciona

Dada um conjunto com a amostra mestre aleatória $x_1, x_2, \ldots, x_n$ de uma população qualquer, devemos sortear consecutivamente, com reposição, conjuntos de tamanho também n dessa amostra amostra mestre. Esses conjuntos são chamados de reamostragem e esse processo deve ser feito milhares de vezes.

A cada reamostragem deve-se aplicar a medida de interesse e guardar em uma lista. Ao fim das milhares de reamostragens, teremos uma distribuição da medida desejada.

De posse desta distribuição, podemos calcular o Intervalo de Confiança, por exemplo, através dos seus percentis.



Atividades...

Pelo Blackboard ou pelo Github, trabalhe com o arquivo:

Atividade 1: Contexto Teórico

`Aula22 Atividade1_Bootstrapping.ipynb`

Atividade 2: Contexto Prático

`Aula22 Atividade2_Bootstrapping.ipynb`

APS6:

ENTREGA DO EXERCÍCIO 3 da Atividade 2 (Em até 3 integrantes para dia 07/05)

Insper