

MINIPROJETO 2

TEOREMA DO LIMITE CENTRAL (TLC)

ESTUDO DA DISTRIBUIÇÃO DA MÉDIA AMOSTRAL (\bar{X}) VIA SIMULAÇÃO

OBJETIVO:

O objetivo deste miniprojeto é que o aluno seja capaz de **compreender e explicar o resultado do Teorema do Limite Central (TLC) por meio de simulação**, uma vez que a demonstração teórica é complexa e requer conhecimento de outras teorias avançadas.

A seguir, apresentamos a definição da média amostral e o resultado descrito no Teorema do Limite Central.

Definição:

Seja X uma variável aleatória de interesse com distribuição qualquer, cuja média populacional e variância populacional são, usualmente, denotadas por $E(X) = \mu$ e $Var(X) = \sigma^2$.

Para estimar a **média populacional**, μ , utiliza-se o estimador **Média Amostral** definido por:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}.$$

Teorema do Limite Central:

Para uma amostra aleatória retirada de uma população cuja variável de interesse X tem média μ e variância σ^2 , então a distribuição de da média amostral (\bar{X}) se aproxima de uma distribuição normal com média μ e variância σ^2/n , quando n tende ao infinito (suficientemente grande).

Parafraseando, seja X uma variável aleatória com média μ e variância σ^2 , então $\bar{X} \sim N\left(\mu; \frac{\sigma^2}{n}\right)$, quando n tende ao infinito (suficientemente grande). Mesmo que X não tenha distribuição normal

O QUE DEVE SER FEITO:

Neste miniprojeto, você e sua equipe (até TRIO) deverão fazer um estudo de **SIMULAÇÃO** para visualizar graficamente a distribuição da média amostral, calcular o valor médio e a variância dessas médias e interpretar esses resultados.

Para tanto, siga os seguintes passos:

1. Construir os valores da população:
 - a. Definir uma distribuição teórica para variável X e definir valor(es) para representar o(s) parâmetro(s) da sua distribuição teórica escolhida.
 - i. Por exemplo, se definir que $X \sim \text{Poisson}(\lambda)$, então, para essa distribuição teórica, tem-se que $E(X) = \text{Var}(X) = \lambda$. Logo, será necessário definir apenas um valor para λ , por exemplo, $\lambda = 4$.
 - b. Sortear aleatoriamente um número muito grande de valores da sua distribuição teórica definida no item 1.a. Esses valores gerados aleatoriamente irão caracterizar todos os valores da sua população. Pesquise o método `.rvs` da sua distribuição no Python.
 - i. Continuando o exemplo acima da Poisson, aqui seria `stats.poisson.rvs(mu=4, loc=0, size=1000000)`.
 - c. Calcule a média e a variância desses valores, os quais irão representar, respectivamente, a média populacional μ e a variância populacional σ^2 .
2. Construir a distribuição da média amostral:
 - a. Inicialmente, considere $n = 2$.
 - b. Sorteie n valores entre os muitos gerados no item 1b. Para tanto, pesquise o método `.random.choice` do `numpy`.
 - c. Calcule a média amostral dos valores obtidos no item anterior.
 - d. Repita os itens 2.b e 2.c por 10.000 vezes. Nesse caso, deverá ter 10.000 médias amostrais.
 - e. Considerando essas 10.000 médias amostrais, construa o histograma com esses valores, calcule a média e a variância dessas médias amostrais.
 - f. Interprete os resultados confrontando com o resultado aguardado via TLC.
3. Repita o item 2 para $n = 4$; $n = 20$; $n = 50$.
4. Faça uma conclusão geral explicando tudo o que aconteceu, o que é para ser visto e como se relaciona com o TLC. Faça essa explicação como se alguém pudesse aprender o TLC a partir de sua simulação.

Ainda, para compreender o resultado do TLC, é necessário observar que a sequência (X_1, \dots, X_n) é chamada de **amostra aleatória**, ou seja, as variáveis dessa sequência são independentes e identicamente distribuídas a X .

Para tanto, siga os seguintes passos:

- i. Considerando as simulações de $n = 50$, por exemplo, guarde sempre o primeiro e o nono valores sorteados em cada uma das 10.000 vezes. Ao final, deverá ter duas listas de 10.000 valores cada, uma que irá representar X_1 e outra que irá representar X_9 .
- ii. Faça o histograma apenas com os valores X_1 .
- iii. Faça o histograma apenas com os valores X_9 .
- iv. Compare com a distribuição de X com que escolheram no item 1 acima. Elas são iguais? Ou seja, X_1 e X_9 são identicamente distribuídas a X ?
- v. Calcule a correlação entre X_1 e X_9 e verifique se essa correlação é próxima de zero, indicando a independência entre elas.

ENTREGÁVEIS ESPERADOS E DATAS:

Turmas A, B e C:

Item	Data	Descrição
Entrega	14/10/2016	Arquivo ipynb enviado na pasta MiniProjeto2 no Github. ATENÇÃO: <ol style="list-style-type: none">1. Todos os integrantes do grupo devem ter o MiniProjeto2 no seu Github.2. Indicar os nomes dos integrantes da equipe logo no topo do conteúdo do arquivo.

RUBRICS DE AVALIAÇÃO DO OBJETIVO DE APRENDIZADO

Objetivo de aprendizado	Insatisfatório (I)	Em desenvolvimento (D)	Essencial (C)	Proficiente (B)	Avançado (A)
Compreender o resultado do TLC via simulação	Apresentou entregas insuficientes ou atrasadas fora do prazo combinado com o professor.	<p>Escolheu um modelo para a variável X considerando valores para a média e a variância dessa variável.</p> <p>A simulação não só é incompleta como também não dá o menor indicativo que seria possível simular o TLC.</p>	Apresenta simulações mas não estão conforme as instruções deste enunciado	Simulou exatamente como no enunciado e explica basicamente o TLC	<p>Simulou exatamente como no enunciado</p> <p>Procurou complementar a simulação com detalhes extra que reforcem a compreensão do TLC (fórmulas, gráficos, etc)</p> <p>Explica o TLC a ponto de se poder estudar pelo seu notebook</p>