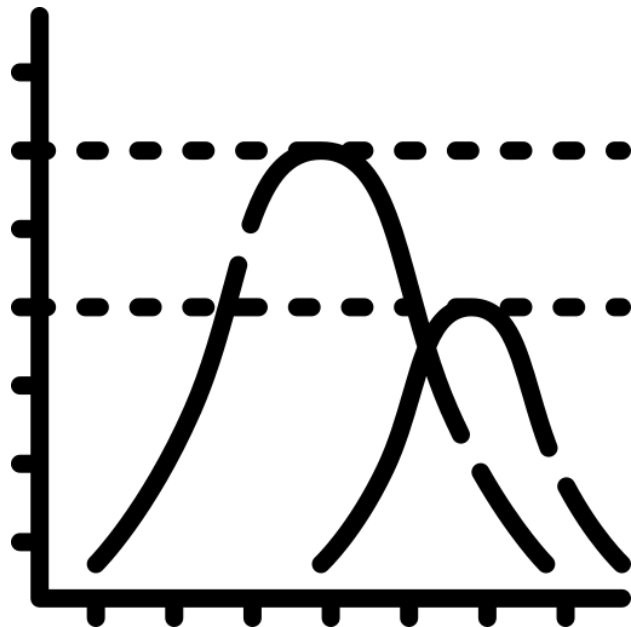


altura



# ESTATÍSTICA COM PYTHON - PARTE 2

AMOSTRAGEM E ESTIMAÇÃO

*Por: Arthur Braz*

## OBJETIVO

- Exercitar conhecimentos de cálculo de probabilidade, amostragem e estimação com os dados da Pesquisa Nacional por Amostra de Domicílios - 2015 do IBGE.

## VARIÁVEIS UTILIZADAS

O conjunto de dados possui 76840 linhas e 7 colunas, das quais as duas que serão utilizadas são:

- Renda: Rendimento mensal do trabalho principal para pessoas de 10 anos ou mais de idade.

## PROBLEMA

Uma nova pesquisa de campo será feita para avaliar o rendimento dos chefes de domicílio no Brasil e com isso algumas condições devem ser satisfeitas:

- Orçamento de 150.000,00 reais.
- Margem de erro não deve ser superior a 10%.

Sabe-se previamente que o custo médio por indivíduo entrevistado fica em torno de 100,00 reais.

Nesse sentido, assumindo um nível de confiança de 95%, qual margem de erro pode ser considerada utilizando todo o recurso disponível?

Para solucionar o problema primeiro preciso encontrar o número máximo de entrevistas que podem ser feitos com o recurso disponível, para isso divido a quantidade de recurso pelo custo por indivíduo entrevistado:

$$Total\ de\ entrevistas = \frac{recurso}{custo\ por\ entrevista} = \frac{150000}{100} = 1500$$

Com isso é possível ser feita até 1500 entrevista com recurso disponível, ou seja o esse é o tamanho máximo da minha amostra. Agora preciso encontrar as estimativas da população, tanto a média e o desvio padrão da amostra serão obtidas através do python e por se tratar da variável Renda ambas estão em reais.

```
#Obtendo 200 amostras de renda do dataset
amostra = dados.Renda.sample(n=200, random_state = 101)
media_amostrai = amostra.mean() # Obtendo a média
desvio_amostrai = amostra.std() # Obtendo o desvio-padrão
```

$$Média\ amostral = \bar{x} = 1964.205$$

$$Desvio\ padrão\ o\ amostral = s = 3139.89$$

Como o nível de confiança é de 95% o score  $z$  é igual a 1,96, com esses resultados já possível encontrar o erro amostral:

$$erro = z \cdot \frac{s}{\sqrt{n}} = 1,96 \cdot \frac{3139,89}{\sqrt{1500}} = 158,90 \text{ reais}$$

Com isso tenho um erro de aproximadamente 159 reais em torno da média populacional, para chegar na margem de erro basta apenas dividir pela estimativa da média.

$$erro \text{ percentual} = \frac{erro}{\bar{x}} \approx 0,0809$$

## Conclusão

Utilizando todo o recurso de 150.000,00 reais é possível fazer uma pesquisa com uma margem de erro de 8,09%.

Ressalto que esse é apenas um dos problema que resolvi no curso de estatística e probabilidade da Alura, se tiver interesse em ver o notebook completo do projeto pode ser acessado no meu [Github](#).