



# Data Science Academy

[www.datascienceacademy.com.br](http://www.datascienceacademy.com.br)

## Microsoft Power BI Para Data Science

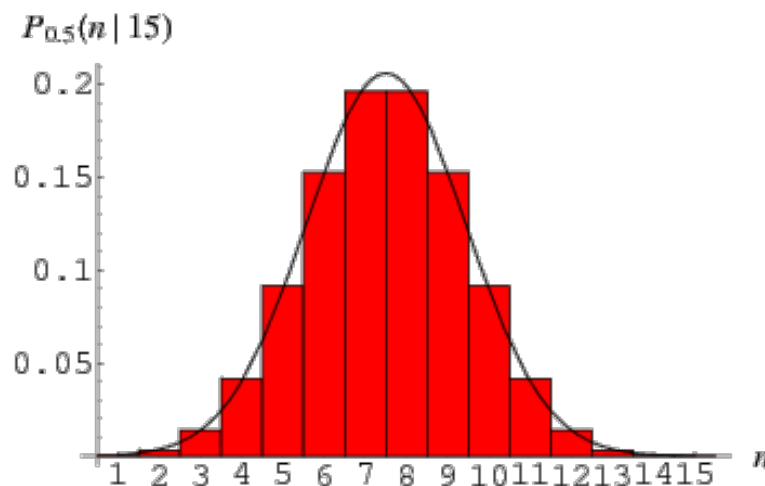
### Teorema do Limite Central

O Teorema do Limite Central é um importante conceito da estatística e a demonstração de muitos outros teoremas estatísticos dependem dele, tanto que é conhecido como a "mãe de todos os teoremas" e realmente merece sua atenção.

Esse teorema afirma que quando o tamanho da amostra aumenta, a distribuição amostral da sua média aproxima-se cada vez mais de uma distribuição normal. Este resultado é fundamental na **teoria da inferência estatística** e sem este teorema, provavelmente a estatística não teria avançado como a ciência que é hoje.

Como já vimos, uma variável aleatória pode ter uma distribuição, possuindo uma média  $\mu$  (**Mu**) e um desvio padrão  $\sigma$  (**Sigma**).

Se, ao invés de tirarmos uma única amostra tirarmos várias amostras de tamanho ' $n$ ' (digamos 20 amostras) e analisarmos a distribuição das médias de cada amostra de tamanho ' $n$ ', observaremos que: à medida que o tamanho ' $n$ ' da amostra aumenta, a distribuição das médias amostrais tende a uma distribuição normal.



Para  $n \geq 30$ , a distribuição das médias amostrais pode ser aproximada satisfatoriamente por uma distribuição normal.

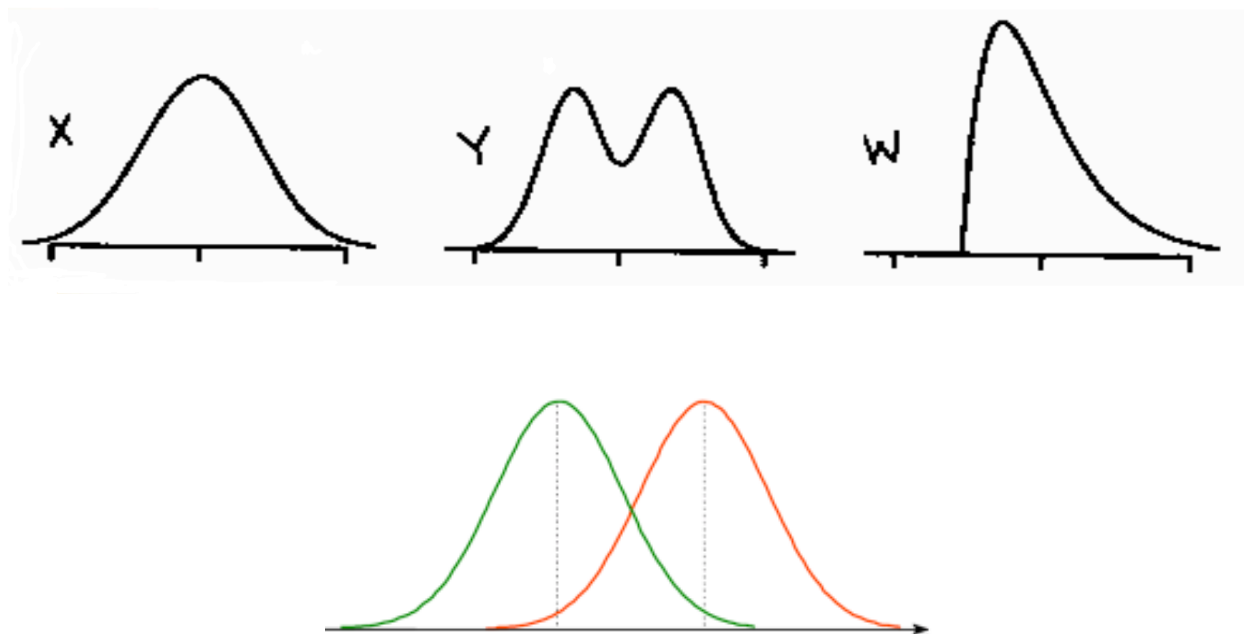
**A média das médias amostrais será a média populacional.**

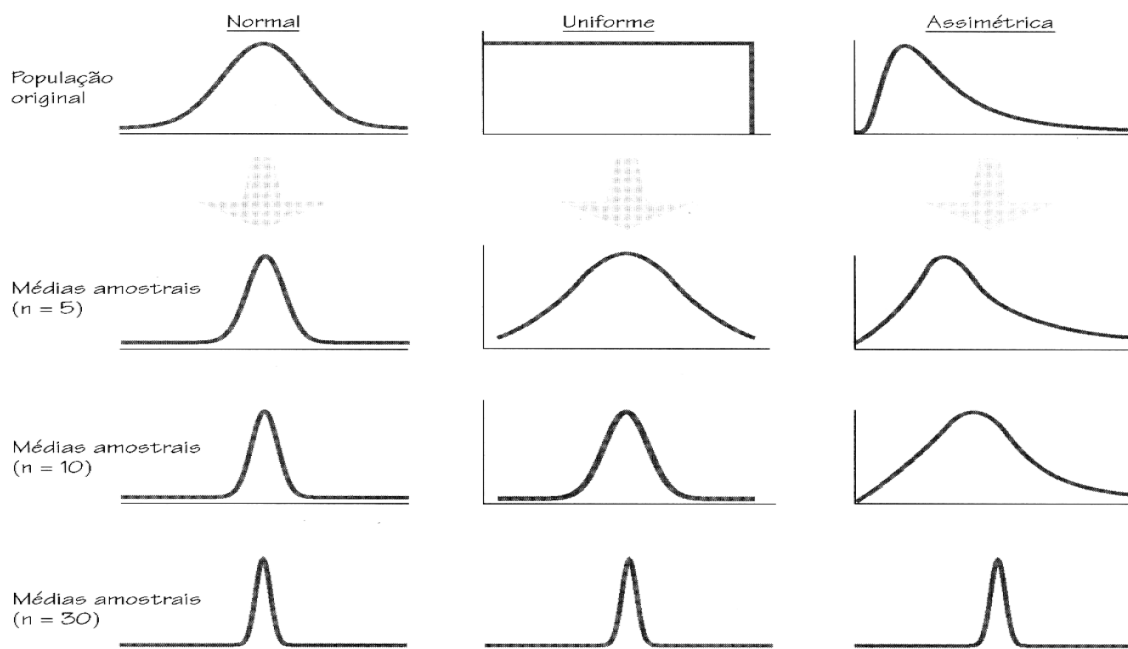
Se a distribuição da variável 'x' for originalmente uma distribuição normal, então a distribuição das médias amostrais terá distribuição normal para qualquer tamanho amostral 'n'. De acordo com o **Teorema do Limite Central**, médias amostrais de amostras suficientemente grandes, retiradas de qualquer população, terão uma distribuição normal.

Se a população seguir uma **distribuição normal** de probabilidade, as **médias amostrais** também terão **distribuição normal**, independente do tamanho das amostras.

Mas afinal, o que há de extraordinário no Teorema do Limite Central? Ele nos diz que qualquer que seja a forma da distribuição original, suas médias resultam em uma distribuição normal. Esse teorema possibilita medir o quanto sua média amostral irá variar, sem ter que pegar outra média amostral para fazer a comparação. **Ou seja, permite-nos conduzir alguns procedimentos de inferência sem ter qualquer conhecimento da distribuição da população.**

Por exemplo, todas as 3 densidades abaixo têm a mesma média e desvio padrão, apesar de suas formas diferentes. Mas as distribuições das médias das amostras serão praticamente idênticas.





Perceba que independente do formato original da população, a distribuição das médias amostrais tende sempre para uma distribuição normal.