# Introdução à estatística

## Aula 4 - Distribuição Normal de Probabilidade

Felipe Nunes, Ph.D.

October 18, 2016

Curso de aperfeiçoamento BR040

# Distribuições Contínuas: Normal

- Dentre as várias distribuições de probabilidade contínuas, vamos concentrar apenas na distribuição normal, pois ela apresenta grande aplicação em pesquisas científicas e tecnológicas.
- Grande parte das variáveis contínuas de interesse prático segue essa distribuição, aliada ao Teorema do Limite Central (TLC), que é a base das estimativas e dos testes de hipóteses realizados sobre a média de uma população qualquer.

# Distribuições Contínuas: Normal

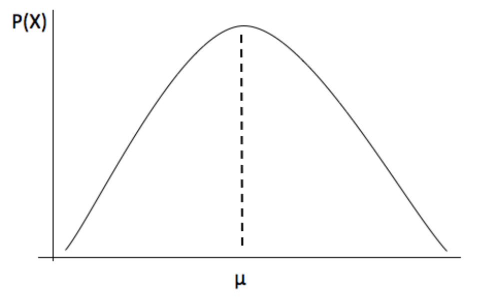
- Vamos tentar desmistificar a idéia de que a variável tem que seguir uma distribuição normal para se usar uma distribuição normal, e explicar o significado de ter a distribuição amostral das médias seguindo uma distribuição normal.
- A distribuição normal de probabilidades é aplicada a variáveis quantitativas contínuas que tem as probabilidades distribuídas de forma simétrica e mesocurdica.

# Distribuições Contínuas: Normal

- Vamos tentar desmistificar a idéia de que a variável tem que seguir uma distribuição normal para se usar uma distribuição normal, e explicar o significado de ter a distribuição amostral das médias seguindo uma distribuição normal.
- Os estimadores amostrais que utilizamos para estimar parâmetros populacionais são normalmente distribuídos, mesmo quando a variável não é. Isso é o que nos permite garantir o tamanho do erro em uma estimativa .

# Distribuições Contínuas: Normal

- A distribuição normal de probabilidades é aplicada a variáveis quantitativas contínuas. O gráfico da distribuição normal é dado por:



# Distribuições Contínuas: Normal

- A distribuição normal de probabilidade tem as seguintes propriedades:

  1. é simétrica em relação ao ponto $x = \mu$ (50% abaixo e 50% acima da média);
  2. as três medidas de posição - média, mediana e moda - se confundem no ponto máximo da curva ($x = \mu$);
  3. fica perfeitamente definida conhecendo-se a média e o desvio padrão, pois outros termos da função são constantes; e
  4. toda a área compreendida entre a curva e o eixo X é igual a 1

# Distribuições Contínuas: Normal

- A função densidade de probabilidade da distribuição normal é dada por:

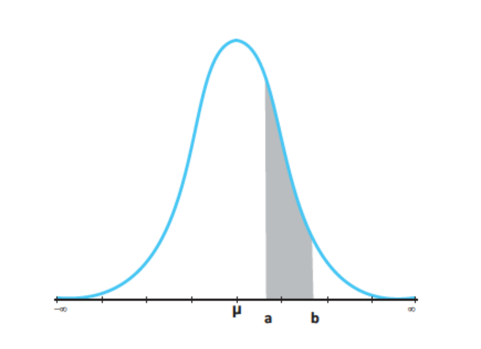
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}.\exp^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

  onde $\mu$ e $\sigma^2$ são a média e a variância, respectivamente, da distribuição de probabilidade; e $\pi$ corresponde a 3,1415 e exp a uma função exponencial.

- Simbolicamente, definimos $X \sim Normal(\mu, \sigma^2)$

# Distribuições Contínuas: Normal

- Portanto, a área sob a curva entre os pontos a e b, em que $a < b$, representa a probabilidade da variável X assumir um valor entre a e b (área escura), como observamos a seguir.



# Distribuições Contínuas: Normal

- Diferentemente do caso de distribuições discretas, para calcularmos as probabilidades via distribuição normal, é necessário o conhecimento de cálculo integral.
- Para evitar isso procuramos tabelar os valores de probabilidade que seriam obtidos por meio da integração da função densidade de probabilidade normal em um determinado intervalo (tabelas de probabilidades).

# Distribuições Contínuas: Normal

- A dificuldade para se processar esse tabelamento se prendeu na infinidade de valores que $\mu$ e $\sigma^2$ poderiam assumir. Nessas condições, teríamos que dispor de uma tabela para cada uma das infinitas combinações de $\mu$ e $\sigma^2$ em cada situação que se quisesse calcular uma probabilidade.
- Para resolver esse problema, convencionou-se que todo o espaço da distribuição normal serial correspondente a 1, e todos os valores da distribuição poderiam ser padronizados em torno do desvio-padrão.

# Distribuições Contínuas: Normal

- O problema foi solucionado mediante emprego de uma nova variável definida por:

$$Z = \frac{x - \mu}{\sigma}$$

- Essa variável transforma todas as distribuições normais em uma distribuição normal reduzida ou padronizada, de média zero e desvio padrão um.
- Então, temos: $Z \sim Normal(0, 1)$

# Distribuições Contínuas: Normal

- Portanto, para um valor de $x = \mu$ em uma distribuição normal qualquer, corresponde o valor:

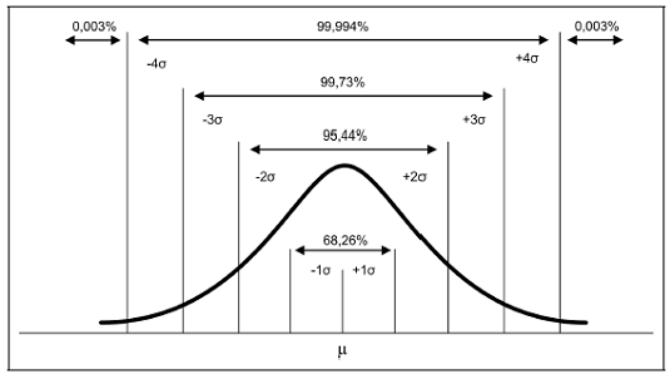
$$Z = \frac{x - \mu}{\sigma} = \frac{\mu - \mu}{\sigma} = 0$$

- Para $x = \mu + \sigma$, temos:

$$Z = \frac{x - \mu}{\sigma} = \frac{\mu + \sigma - \mu}{\sigma} = \frac{\sigma}{\sigma} = 1$$

# Distribuições Contínuas: Normal

- Por conta dessa padronização possível, uma distribuição normal padronizada nos possibilita identificar:



# Distribuições Contínuas: Normal

- Ao padronizar uma distribuição normal é possível usar apenas uma tabela de probabilidades que independente dos valores de $\mu$ e $\sigma^2$.
- Vamos usar como referência uma tabela em que os valores de áreas ou probabilidades fornecidas estão **entre zero e o valor de Z** (ver tabela e fazer um exemplo gráfico).
- Na tabela de probabilidades os valores apresentados na primeira coluna correspondem a parte inteira e decimal do valor de Z, enquanto os valores da primeira linha correspondem a parte centesimal.

# Distribuições Contínuas: Normal

| z | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 0.0000 | 0.0040 | 0.0080 | 0.0120 | 0.0160 | 0.0199 | 0.0239 | 0.0279 | 0.0319 | 0.0359 |
| 0.1 | 0.0398 | 0.0438 | 0.0478 | 0.0517 | 0.0557 | 0.0596 | 0.0636 | 0.0675 | 0.0714 | 0.0753 |
| 0.2 | 0.0793 | 0.0832 | 0.0871 | 0.0910 | 0.0948 | 0.0987 | 0.1026 | 0.1064 | 0.1103 | 0.1141 |
| 0.3 | 0.1179 | 0.1217 | 0.1255 | 0.1293 | 0.1331 | 0.1368 | 0.1406 | 0.1443 | 0.1480 | 0.1517 |
| 0.4 | 0.1554 | 0.1591 | 0.1628 | 0.1664 | 0.1700 | 0.1736 | 0.1772 | 0.1808 | 0.1844 | 0.1879 |
| 0.5 | 0.1915 | 0.1950 | 0.1985 | 0.2019 | 0.2054 | 0.2088 | 0.2123 | 0.2157 | 0.2190 | 0.2224 |
| 0.6 | 0.2257 | 0.2291 | 0.2324 | 0.2357 | 0.2389 | 0.2422 | 0.2454 | 0.2486 | 0.2517 | 0.2549 |
| 0.7 | 0.2580 | 0.2611 | 0.2642 | 0.2673 | 0.2704 | 0.2734 | 0.2764 | 0.2794 | 0.2823 | 0.2852 |
| 0.8 | 0.2881 | 0.2910 | 0.2939 | 0.2967 | 0.2995 | 0.3023 | 0.3051 | 0.3078 | 0.3106 | 0.3133 |
| 0.9 | 0.3159 | 0.3186 | 0.3212 | 0.3238 | 0.3264 | 0.3289 | 0.3315 | 0.3340 | 0.3365 | 0.3389 |
| 1.0 | 0.3413 | 0.3438 | 0.3461 | 0.3485 | 0.3508 | 0.3531 | 0.3554 | 0.3577 | 0.3599 | 0.3621 |
| 1.1 | 0.3643 | 0.3665 | 0.3686 | 0.3708 | 0.3729 | 0.3749 | 0.3770 | 0.3790 | 0.3810 | 0.3830 |
| 1.2 | 0.3849 | 0.3869 | 0.3888 | 0.3907 | 0.3925 | 0.3944 | 0.3962 | 0.3980 | 0.3997 | 0.4015 |
| 1.3 | 0.4032 | 0.4049 | 0.4066 | 0.4082 | 0.4099 | 0.4115 | 0.4131 | 0.4147 | 0.4162 | 0.4177 |
| 1.4 | 0.4192 | 0.4207 | 0.4222 | 0.4236 | 0.4251 | 0.4265 | 0.4279 | 0.4292 | 0.4306 | 0.4319 |
| 1.5 | 0.4332 | 0.4345 | 0.4357 | 0.4370 | 0.4382 | 0.4394 | 0.4406 | 0.4418 | 0.4429 | 0.4441 |
| 1.6 | 0.4452 | 0.4463 | 0.4474 | 0.4484 | 0.4495 | 0.4505 | 0.4515 | 0.4525 | 0.4535 | 0.4545 |
| 1.7 | 0.4554 | 0.4564 | 0.4573 | 0.4582 | 0.4591 | 0.4599 | 0.4608 | 0.4616 | 0.4625 | 0.4633 |
| 1.8 | 0.4641 | 0.4649 | 0.4656 | 0.4664 | 0.4671 | 0.4678 | 0.4686 | 0.4693 | 0.4699 | 0.4706 |
| 1.9 | 0.4713 | 0.4719 | 0.4726 | 0.4732 | 0.4738 | 0.4744 | 0.4750 | 0.4756 | 0.4761 | 0.4767 |
| 2.0 | 0.4772 | 0.4778 | 0.4783 | 0.4788 | 0.4793 | 0.4798 | 0.4803 | 0.4808 | 0.4812 | 0.4817 |
| 2.1 | 0.4821 | 0.4826 | 0.4830 | 0.4834 | 0.4838 | 0.4842 | 0.4846 | 0.4850 | 0.4854 | 0.4857 |
| 2.2 | 0.4861 | 0.4864 | 0.4868 | 0.4871 | 0.4875 | 0.4878 | 0.4881 | 0.4884 | 0.4887 | 0.4890 |
| 2.3 | 0.4893 | 0.4896 | 0.4898 | 0.4901 | 0.4904 | 0.4906 | 0.4909 | 0.4911 | 0.4913 | 0.4916 |
| 2.4 | 0.4918 | 0.4920 | 0.4922 | 0.4925 | 0.4927 | 0.4929 | 0.4931 | 0.4932 | 0.4934 | 0.4936 |

# Distribuições Contínuas: Normal

- **Exemplo**: Vamos considerar a arrecadação com o um tributo de uma pequena cidade. Verificamos que essa arrecadação seguia uma distribuição normal com média de R$ 60.000,00 e desvio padrão de R$ 10.000,00. Procuramos, então, responder os seguintes questionamentos:

  1. Qual a prob. de uma arrecadação ser maior que R$ 75.000?
  2. Qual a prob. da arrecadação estar entre R$ 50.000,00 e R$ 70.000,00?
  3. Qual a probabilidade da arrecadação estar entre R$ 63.000,00 e R$ 70.000,00?

1. Qual a prob. de uma arrecadação ser maior que R$ 75.000?

   - Como a variável arrecadação apresenta distribuição aproximadamente normal com média 60.000 e variância de 10.000, definimos

$$X \sim N(60.000, 10.000)$$

   e procura- se calcular a

$$P(X > 75.000) = ?$$

# Distribuições Contínuas: Normal

- Primeiramente, precisamos transformar a variável X em Z e, depois, substituindo na expressão, teremos:

$$Z = \frac{x - \mu}{\sigma} = \frac{75.000 - 60.000}{10.000} = 1.50$$

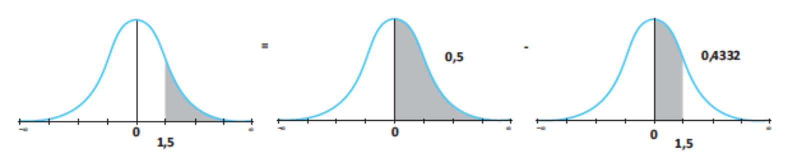
- Olhando esse valor na tabela de probabilidades referente a z = 1,50, encontraremos o valor de 0,4332 que corresponde à probabilidade de $z$ estar entre zero e 1,5.

# Distribuições Contínuas: Normal

| z | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 0.0000 | 0.0040 | 0.0080 | 0.0120 | 0.0160 | 0.0199 | 0.0239 | 0.0279 | 0.0319 | 0.0359 |
| 0.1 | 0.0398 | 0.0438 | 0.0478 | 0.0517 | 0.0557 | 0.0596 | 0.0636 | 0.0675 | 0.0714 | 0.0753 |
| 0.2 | 0.0793 | 0.0832 | 0.0871 | 0.0910 | 0.0948 | 0.0987 | 0.1026 | 0.1064 | 0.1103 | 0.1141 |
| 0.3 | 0.1179 | 0.1217 | 0.1255 | 0.1293 | 0.1331 | 0.1368 | 0.1406 | 0.1443 | 0.1480 | 0.1517 |
| 0.4 | 0.1554 | 0.1591 | 0.1628 | 0.1664 | 0.1700 | 0.1736 | 0.1772 | 0.1808 | 0.1844 | 0.1879 |
| 0.5 | 0.1915 | 0.1950 | 0.1985 | 0.2019 | 0.2054 | 0.2088 | 0.2123 | 0.2157 | 0.2190 | 0.2224 |
| 0.6 | 0.2257 | 0.2291 | 0.2324 | 0.2357 | 0.2389 | 0.2422 | 0.2454 | 0.2486 | 0.2517 | 0.2549 |
| 0.7 | 0.2580 | 0.2611 | 0.2642 | 0.2673 | 0.2704 | 0.2734 | 0.2764 | 0.2794 | 0.2823 | 0.2852 |
| 0.8 | 0.2881 | 0.2910 | 0.2939 | 0.2967 | 0.2995 | 0.3023 | 0.3051 | 0.3078 | 0.3106 | 0.3133 |
| 0.9 | 0.3159 | 0.3186 | 0.3212 | 0.3238 | 0.3264 | 0.3289 | 0.3315 | 0.3340 | 0.3365 | 0.3389 |
| 1.0 | 0.3413 | 0.3438 | 0.3461 | 0.3485 | 0.3508 | 0.3531 | 0.3554 | 0.3577 | 0.3599 | 0.3621 |
| 1.1 | 0.3643 | 0.3665 | 0.3686 | 0.3708 | 0.3729 | 0.3749 | 0.3770 | 0.3790 | 0.3810 | 0.3830 |
| 1.2 | 0.3849 | 0.3869 | 0.3888 | 0.3907 | 0.3925 | 0.3944 | 0.3962 | 0.3980 | 0.3997 | 0.4015 |
| 1.3 | 0.4032 | 0.4049 | 0.4066 | 0.4082 | 0.4099 | 0.4115 | 0.4131 | 0.4147 | 0.4162 | 0.4177 |
| 1.4 | 0.4192 | 0.4207 | 0.4222 | 0.4236 | 0.4251 | 0.4265 | 0.4279 | 0.4292 | 0.4306 | 0.4319 |
| 1.5 | 0.4332 | 0.4345 | 0.4357 | 0.4370 | 0.4382 | 0.4394 | 0.4406 | 0.4418 | 0.4429 | 0.4441 |
| 1.6 | 0.4452 | 0.4463 | 0.4474 | 0.4484 | 0.4495 | 0.4505 | 0.4515 | 0.4525 | 0.4535 | 0.4545 |
| 1.7 | 0.4554 | 0.4564 | 0.4573 | 0.4582 | 0.4591 | 0.4599 | 0.4608 | 0.4616 | 0.4625 | 0.4633 |
| 1.8 | 0.4641 | 0.4649 | 0.4656 | 0.4664 | 0.4671 | 0.4678 | 0.4686 | 0.4693 | 0.4699 | 0.4706 |
| 1.9 | 0.4713 | 0.4719 | 0.4726 | 0.4732 | 0.4738 | 0.4744 | 0.4750 | 0.4756 | 0.4761 | 0.4767 |
| 2.0 | 0.4772 | 0.4778 | 0.4783 | 0.4788 | 0.4793 | 0.4798 | 0.4803 | 0.4808 | 0.4812 | 0.4817 |
| 2.1 | 0.4821 | 0.4826 | 0.4830 | 0.4834 | 0.4838 | 0.4842 | 0.4846 | 0.4850 | 0.4854 | 0.4857 |
| 2.2 | 0.4861 | 0.4864 | 0.4868 | 0.4871 | 0.4875 | 0.4878 | 0.4881 | 0.4884 | 0.4887 | 0.4890 |
| 2.3 | 0.4893 | 0.4896 | 0.4898 | 0.4901 | 0.4904 | 0.4906 | 0.4909 | 0.4911 | 0.4913 | 0.4916 |
| 2.4 | 0.4918 | 0.4920 | 0.4922 | 0.4925 | 0.4927 | 0.4929 | 0.4931 | 0.4932 | 0.4934 | 0.4936 |

# Distribuições Contínuas: Normal

- Em representação gráfica, temos:




- Portanto, $P(X > 75.000) = 0.5 - 0.4332 = 0.0668$

# Distribuições Contínuas: Normal

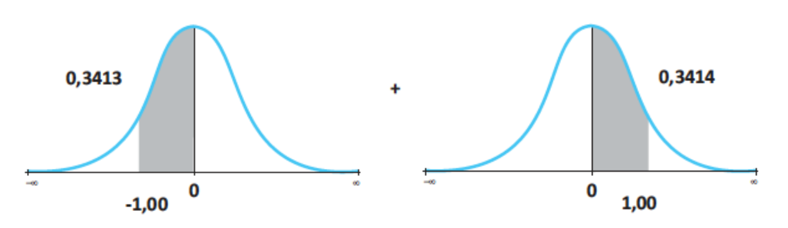
2. Qual a prob. da arrecadação entre R$ 50K e R$ 70K?

   - Estamos buscando $P(50000 < X < 70000) =?$
   - Primeiramente, precisamos transformar a variável X em Z e, depois, substituindo na expressão de Z, teremos valores de $Z_1$ e $Z_2$, relacionados aos valores de $X_1 = 50.000$ e $X_2 = 70.000$

# Distribuições Contínuas: Normal

$$z_1 = \frac{x_1 - \mu}{\sigma_1} = \frac{50.000 - 60.000}{10.000} = -1.00$$

$$z_2 = \frac{x_2 - \mu}{\sigma_1} = \frac{70.000 - 60.000}{10.000} = 1.00$$



# Distribuições Contínuas: Normal

| z | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 0.0000 | 0.0040 | 0.0080 | 0.0120 | 0.0160 | 0.0199 | 0.0239 | 0.0279 | 0.0319 | 0.0359 |
| 0.1 | 0.0398 | 0.0438 | 0.0478 | 0.0517 | 0.0557 | 0.0596 | 0.0636 | 0.0675 | 0.0714 | 0.0753 |
| 0.2 | 0.0793 | 0.0832 | 0.0871 | 0.0910 | 0.0948 | 0.0987 | 0.1026 | 0.1064 | 0.1103 | 0.1141 |
| 0.3 | 0.1179 | 0.1217 | 0.1255 | 0.1293 | 0.1331 | 0.1368 | 0.1406 | 0.1443 | 0.1480 | 0.1517 |
| 0.4 | 0.1554 | 0.1591 | 0.1628 | 0.1664 | 0.1700 | 0.1736 | 0.1772 | 0.1808 | 0.1844 | 0.1879 |
| 0.5 | 0.1915 | 0.1950 | 0.1985 | 0.2019 | 0.2054 | 0.2088 | 0.2123 | 0.2157 | 0.2190 | 0.2224 |
| 0.6 | 0.2257 | 0.2291 | 0.2324 | 0.2357 | 0.2389 | 0.2422 | 0.2454 | 0.2486 | 0.2517 | 0.2549 |
| 0.7 | 0.2580 | 0.2611 | 0.2642 | 0.2673 | 0.2704 | 0.2734 | 0.2764 | 0.2794 | 0.2823 | 0.2852 |
| 0.8 | 0.2881 | 0.2910 | 0.2939 | 0.2967 | 0.2995 | 0.3023 | 0.3051 | 0.3078 | 0.3106 | 0.3133 |
| 0.9 | 0.3159 | 0.3186 | 0.3212 | 0.3238 | 0.3264 | 0.3289 | 0.3315 | 0.3340 | 0.3365 | 0.3389 |
| 1.0 | 0.3413 | 0.3438 | 0.3461 | 0.3485 | 0.3508 | 0.3531 | 0.3554 | 0.3577 | 0.3599 | 0.3621 |
| 1.1 | 0.3643 | 0.3665 | 0.3686 | 0.3708 | 0.3729 | 0.3749 | 0.3770 | 0.3790 | 0.3810 | 0.3830 |
| 1.2 | 0.3849 | 0.3869 | 0.3888 | 0.3907 | 0.3925 | 0.3944 | 0.3962 | 0.3980 | 0.3997 | 0.4015 |
| 1.3 | 0.4032 | 0.4049 | 0.4066 | 0.4082 | 0.4099 | 0.4115 | 0.4131 | 0.4147 | 0.4162 | 0.4177 |
| 1.4 | 0.4192 | 0.4207 | 0.4222 | 0.4236 | 0.4251 | 0.4265 | 0.4279 | 0.4292 | 0.4306 | 0.4319 |
| 1.5 | 0.4332 | 0.4345 | 0.4357 | 0.4370 | 0.4382 | 0.4394 | 0.4406 | 0.4418 | 0.4429 | 0.4441 |
| 1.6 | 0.4452 | 0.4463 | 0.4474 | 0.4484 | 0.4495 | 0.4505 | 0.4515 | 0.4525 | 0.4535 | 0.4545 |
| 1.7 | 0.4554 | 0.4564 | 0.4573 | 0.4582 | 0.4591 | 0.4599 | 0.4608 | 0.4616 | 0.4625 | 0.4633 |
| 1.8 | 0.4641 | 0.4649 | 0.4656 | 0.4664 | 0.4671 | 0.4678 | 0.4686 | 0.4693 | 0.4699 | 0.4706 |
| 1.9 | 0.4713 | 0.4719 | 0.4726 | 0.4732 | 0.4738 | 0.4744 | 0.4750 | 0.4756 | 0.4761 | 0.4767 |
| 2.0 | 0.4772 | 0.4778 | 0.4783 | 0.4788 | 0.4793 | 0.4798 | 0.4803 | 0.4808 | 0.4812 | 0.4817 |
| 2.1 | 0.4821 | 0.4826 | 0.4830 | 0.4834 | 0.4838 | 0.4842 | 0.4846 | 0.4850 | 0.4854 | 0.4857 |
| 2.2 | 0.4861 | 0.4864 | 0.4868 | 0.4871 | 0.4875 | 0.4878 | 0.4881 | 0.4884 | 0.4887 | 0.4890 |
| 2.3 | 0.4893 | 0.4896 | 0.4898 | 0.4901 | 0.4904 | 0.4906 | 0.4909 | 0.4911 | 0.4913 | 0.4916 |
| 2.4 | 0.4918 | 0.4920 | 0.4922 | 0.4925 | 0.4927 | 0.4929 | 0.4931 | 0.4932 | 0.4934 | 0.4936 |

- Podemos verificar que:

$$
\begin{aligned}
P(50000 < X < 70000) &= P(-1,00 < z < 1,00) \\
&= 0,3413 + 0,3413 \\
&= 0,6826
\end{aligned}
$$

# Distribuições Contínuas: Normal

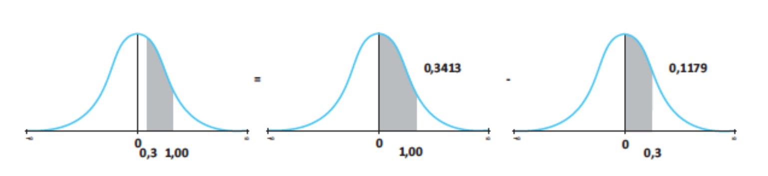
2. Qual a prob. da arrecadação estar entre R$ 63K e R$ 70K?

- Estamos buscando $P(63000 < X < 70000) =?$

$$z_1 = \frac{x_1 - \mu}{\sigma_1} = \frac{63.000 - 60.000}{10.000} = 0.30$$

$$z_2 = \frac{x_2 - \mu}{\sigma_1} = \frac{70.000 - 60.000}{10.000} = 1.00$$

# Distribuições Contínuas: Normal



## Distribuições Contínuas: Normal

- Podemos verificar que:

$$
\begin{aligned}
P(63.000 < X < 70.000) &= P(0.30 < z < 1,00) \\
&= 0,3413 - 0,1179 \\
&= 0,2234
\end{aligned}
$$

# Distribuições Contínuas: Normal

| z | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 0.0000 | 0.0040 | 0.0080 | 0.0120 | 0.0160 | 0.0199 | 0.0239 | 0.0279 | 0.0319 | 0.0359 |
| 0.1 | 0.0398 | 0.0438 | 0.0478 | 0.0517 | 0.0557 | 0.0596 | 0.0636 | 0.0675 | 0.0714 | 0.0753 |
| 0.2 | 0.0793 | 0.0832 | 0.0871 | 0.0910 | 0.0948 | 0.0987 | 0.1026 | 0.1064 | 0.1103 | 0.1141 |
| 0.3 | 0.1179 | 0.1217 | 0.1255 | 0.1293 | 0.1331 | 0.1368 | 0.1406 | 0.1443 | 0.1480 | 0.1517 |
| 0.4 | 0.1554 | 0.1591 | 0.1628 | 0.1664 | 0.1700 | 0.1736 | 0.1772 | 0.1808 | 0.1844 | 0.1879 |
| 0.5 | 0.1915 | 0.1950 | 0.1985 | 0.2019 | 0.2054 | 0.2088 | 0.2123 | 0.2157 | 0.2190 | 0.2224 |
| 0.6 | 0.2257 | 0.2291 | 0.2324 | 0.2357 | 0.2389 | 0.2422 | 0.2454 | 0.2486 | 0.2517 | 0.2549 |
| 0.7 | 0.2580 | 0.2611 | 0.2642 | 0.2673 | 0.2704 | 0.2734 | 0.2764 | 0.2794 | 0.2823 | 0.2852 |
| 0.8 | 0.2881 | 0.2910 | 0.2939 | 0.2967 | 0.2995 | 0.3023 | 0.3051 | 0.3078 | 0.3106 | 0.3133 |
| 0.9 | 0.3159 | 0.3186 | 0.3212 | 0.3238 | 0.3264 | 0.3289 | 0.3315 | 0.3340 | 0.3365 | 0.3389 |
| 1.0 | 0.3413 | 0.3438 | 0.3461 | 0.3485 | 0.3508 | 0.3531 | 0.3554 | 0.3577 | 0.3599 | 0.3621 |
| 1.1 | 0.3643 | 0.3665 | 0.3686 | 0.3708 | 0.3729 | 0.3749 | 0.3770 | 0.3790 | 0.3810 | 0.3830 |
| 1.2 | 0.3849 | 0.3869 | 0.3888 | 0.3907 | 0.3925 | 0.3944 | 0.3962 | 0.3980 | 0.3997 | 0.4015 |
| 1.3 | 0.4032 | 0.4049 | 0.4066 | 0.4082 | 0.4099 | 0.4115 | 0.4131 | 0.4147 | 0.4162 | 0.4177 |
| 1.4 | 0.4192 | 0.4207 | 0.4222 | 0.4236 | 0.4251 | 0.4265 | 0.4279 | 0.4292 | 0.4306 | 0.4319 |
| 1.5 | 0.4332 | 0.4345 | 0.4357 | 0.4370 | 0.4382 | 0.4394 | 0.4406 | 0.4418 | 0.4429 | 0.4441 |
| 1.6 | 0.4452 | 0.4463 | 0.4474 | 0.4484 | 0.4495 | 0.4505 | 0.4515 | 0.4525 | 0.4535 | 0.4545 |
| 1.7 | 0.4554 | 0.4564 | 0.4573 | 0.4582 | 0.4591 | 0.4599 | 0.4608 | 0.4616 | 0.4625 | 0.4633 |
| 1.8 | 0.4641 | 0.4649 | 0.4656 | 0.4664 | 0.4671 | 0.4678 | 0.4686 | 0.4693 | 0.4699 | 0.4706 |
| 1.9 | 0.4713 | 0.4719 | 0.4726 | 0.4732 | 0.4738 | 0.4744 | 0.4750 | 0.4756 | 0.4761 | 0.4767 |
| 2.0 | 0.4772 | 0.4778 | 0.4783 | 0.4788 | 0.4793 | 0.4798 | 0.4803 | 0.4808 | 0.4812 | 0.4817 |
| 2.1 | 0.4821 | 0.4826 | 0.4830 | 0.4834 | 0.4838 | 0.4842 | 0.4846 | 0.4850 | 0.4854 | 0.4857 |
| 2.2 | 0.4861 | 0.4864 | 0.4868 | 0.4871 | 0.4875 | 0.4878 | 0.4881 | 0.4884 | 0.4887 | 0.4890 |
| 2.3 | 0.4893 | 0.4896 | 0.4898 | 0.4901 | 0.4904 | 0.4906 | 0.4909 | 0.4911 | 0.4913 | 0.4916 |
| 2.4 | 0.4918 | 0.4920 | 0.4922 | 0.4925 | 0.4927 | 0.4929 | 0.4931 | 0.4932 | 0.4934 | 0.4936 |

# O que tem que ser normal afinal?

# O que tem que ser normal afinal?

- Imagine uma população de alunos com 4 componentes.
- Eles tem as seguintes notas na prova final:

$$X1 = 40; X2 = 60; X3 = 80; X4 = 100$$

- Qual a média das notas dos alunos?

$$\mu = \frac{\sum x_i}{n} = \frac{40 + 60 + 80 + 100}{4} = \frac{280}{4} = 70$$

# O que tem que ser normal afinal?

- Essa variável segue uma distribuição normal?

| X | Freq | P(X) |
|---|---|---|
| 40 | 1 | 1/4 |
| 60 | 1 | 1/4 |
| 80 | 1 | 1/4 |
| 100 | 1 | 1/4 |

- NÃO!!! Ela segue uma distribuição Uniforme:

$$X \sim Uniforme(a, b)$$

## O que tem que ser normal afinal?

- Mas e as possíveis médias de X dessa população ($n = 4$) com amostras tamanho $n = 2$? (Vamos listar todas)

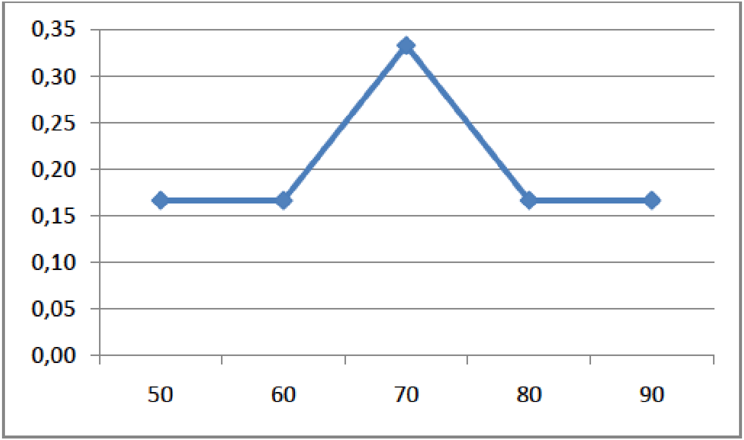
$$
\begin{aligned}
\bar{x}_1 &= \frac{x_1 + x_2}{2} = (40 + 60)/2 = 50 \\
\bar{x}_2 &= \frac{x_1 + x_3}{2} = (40 + 80)/2 = 60 \\
\bar{x}_3 &= \frac{x_1 + x_4}{2} = (40 + 100)/2 = 70 \\
\bar{x}_4 &= \frac{x_2 + x_3}{2} = (60 + 80)/2 = 70 \\
\bar{x}_5 &= \frac{x_2 + x_4}{2} = (60 + 100)/2 = 80 \\
\bar{x}_6 &= \frac{x_3 + x_4}{2} = (80 + 100)/2 = 90
\end{aligned}
$$

# O que tem que ser normal afinal?

- Vejamos então a probabilidade de ocorrência de cada média amostral:

| X | Freq | P(X) |
|---|---|---|
| 50 | 1 | 1/6 |
| 60 | 1 | 1/6 |
| 70 | 2 | 1/3 |
| 80 | 1 | 1/6 |
| 90 | 1 | 1/6 |

# O que tem que ser normal afinal?



# O que tem que ser normal afinal?

- A distribuição das probabilidades das **possíveis** médias amostrais é quase normal.
- Só não tivemos uma distribuição normal das probabilidades das médias amostrais porque o nosso $n < 30$.
- Vale a regra: **A distribuição das médias amostrai será sempre simétrica, e suas probabilidades serão normalmente distribuídas quando o** $n > 30$.

# Teoria do Limite Central

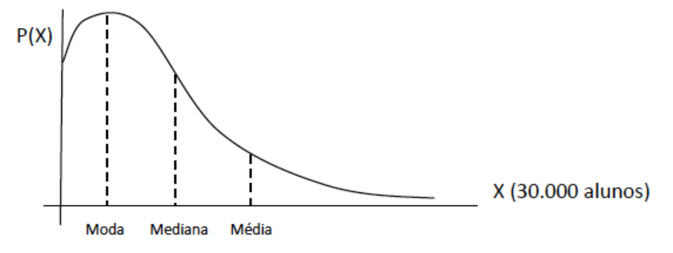
# Teoria do Limite Central

- **TLC**: Independente da forma como as variáveis estão distribuídas na população, a distribuição amostral das possíveis médias, quando a amostra é maior do que 30 ($n > 30$), sempre será normalmente distribuída.

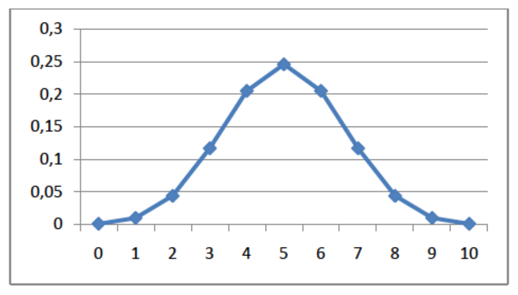
## Teoria do Limite Central

- Podemos fazer um exemplo usando a quantidade de dinheiro que cada um de vocês tem na carteira neste momento.
- Querem fazer pra ver?
- Ou podemos ver um exemplo real!

- **Exemplo**: A UFMG fez uma pesquisa com seus 30.000 alunos para saber a renda per capita da universidade. Tivemos o seguinte resultado:



- Se fizermos n = 1.000 amostras dessa população de N = 30.000 alunos e calcularmos as probabilidades das médias das possíveis amostras teremos o seguinte gráfico de probabilidades das médias amostrais:



# Teoria do Limite Central

- Ou seja, usamos a distribuição normal para fazer estimações populacionais porque a distribuição das probabilidades das médias amostrais é normal, quando a amostra é maior do que 30, independentemente da forma como a variável se comporta na população.
- Vimos que a renda se distribui assimetricamente na população, mas as probabilidades das médias amostrais da renda se distribuem normalmente.

# Propriedade de estimadores

# Propriedade de estimadores

1ª **propriedade dos estimadores**: posso dizer que a média amostral é um estimador **não-tendencioso** porque posso estimar o parâmetro populacional com ele, e o valor esperado da média amostral será igual ao da média populacional.

Ou seja, posso usar a média de uma amostra como uma boa estimativa da média da população, porque nós sabemos que a média das possíveis médias de uma amostra é igual a média populacional, ou

$$E(X) = \sum x.P(x) = \mu$$

# Propriedade de estimadores

Notas da população N = 4:

$$X1 = 40; X2 = 60; X3 = 80; X4 = 100$$

Média populacional ($\mu$):

$$\mu = \frac{\sum x_i}{n} = \frac{40 + 60 + 80 + 100}{4} = \frac{280}{4} = 70$$

Notas médias de todas as possíveis amostras n = 2:

$$
\begin{aligned}
\bar{x_1} &= \frac{x_1 + x_2}{2} = (40 + 60)/2 = 50 \\
\bar{x_2} &= \frac{x_1 + x_3}{2} = (40 + 80)/2 = 60 \\
\bar{x_3} &= \frac{x_1 + x_4}{2} = (40 + 100)/2 = 70 \\
\bar{x_4} &= \frac{x_2 + x_3}{2} = (60 + 80)/2 = 70 \\
\bar{x_5} &= \frac{x_2 + x_4}{2} = (60 + 100)/2 = 80 \\
\bar{x_6} &= \frac{x_3 + x_4}{2} = (80 + 100)/2 = 90
\end{aligned}
$$

Distribuição de prob. de todas as possíveis amostras n = 2:

| X | Freq | P(X) |
|---|---|---|
| 50 | 1 | 1/6 |
| 60 | 1 | 1/6 |
| 70 | 2 | 1/3 |
| 80 | 1 | 1/6 |
| 90 | 1 | 1/6 |

$$E(X) = \sum X.P(X) = \frac{50}{6} + \frac{60}{6} + \frac{70}{3} + \frac{80}{6} + \frac{90}{6} = 70$$

# Propriedade de estimadores

Portanto, a média das possíveis médias de uma amostra é igual a média populacional, ou

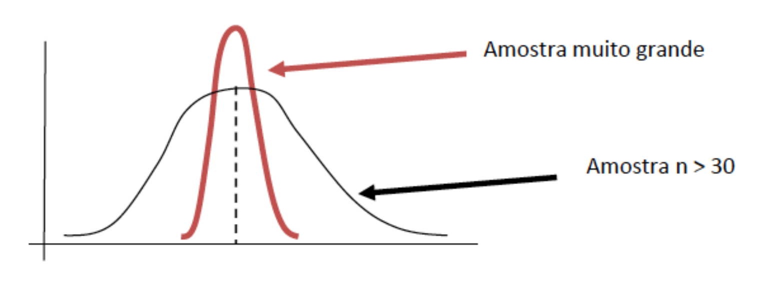
$$E(X) = \mu$$

2ª **propriedade dos estimadores**: propriedade da consistência

Um estimador é consistente se ao aumentar o número de casos na amostra diminui-se o erro de estimação.

Ou seja, quanto maior a minha amostra menor a chance de errar a minha estimação populacional pois a minha curva vai se aproximando cada vez mais da média.

# Propriedade de estimadores



# Resumindo

## Resumindo

- Distribuição Contínua de Probabilidade: Normal
- O que tem que ser normal afinal?
- Teoria do Limite Central
- Propriedade dos estimadores