

Universidade Federal do Ceará Faculdade de Economia

Métodos Quantitativos

Vicente Lima Crisóstomo

Fortaleza, 2020

Sumário

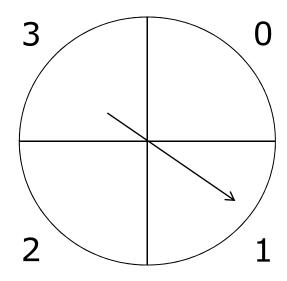
- Introdução
- Estatística Descritiva
- Probabilidade
- Distribuições de Probabilidades
- Amostragem e Distribuições Amostrais
- Estimação
- Testes de Significância
- Análise de Variância
- Teste de Significância para Proporções
- Testes Não Paramétricos
- Correlação e Regressão

Distribuições *Contínuas* de Probabilidades

- Distribuições para variáveis aleatórias
 - Contínuas
 - Discretas com grande número de resultados possíveis
- Distribuição UNIFORME
- Distribuição NORMAL
- Distribuição EXPONENCIAL

Distribuições *Contínuas* de Probabilidades

- Experimentos
 - Com igual probabilidade de cada ocorrência
 - Grande número de ocorrências possíveis

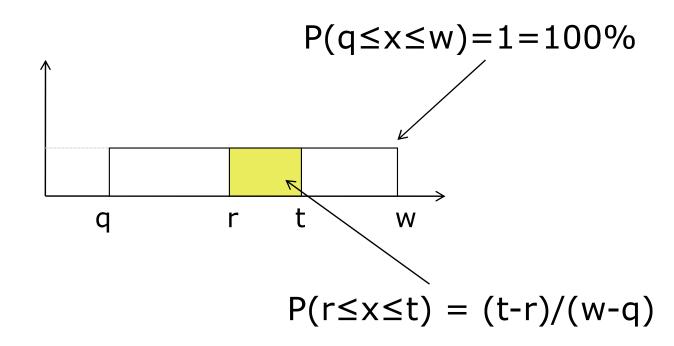


Distribuições *Contínuas* de Probabilidades

- Giro aleatório do ponteiro
 - Círculo pode ser dividido em N partes
 - N = 100 ou 1.000 ou 10.000 ou 100.000 ou 1.000.000
 - P(determinado valor) zero (aproximadamente)
 - Se 10.000.000 de valores/divisões/setores circulares
 - P(1) = P(200) = P(9.999) = 1 / 10.000.000 = 0,0000001 = 0,00001%
 - Se quatro divisões ou setores circulares
 - P(Setor Circular 0) = P(S1) = P(S2) = P(S3) = 1/4 = 25%
 - Se oito divisões ou setores circulares
 - P(cada setor) = 1 / 8 = 0,125 = 12,5%

Distribuição UNIFORME

- VA pode assumir valores igualmente prováveis em um intervalo
- Probabilidades associadas à VA podem ser descritas por uma distribuição Uniforme



Distribuição UNIFORME

Exemplo:

Uma microempresa corta e vende pedaços de madeira cujo comprimento varia entre 50 e 80cm. Qual a probabilidade de um pedaço ter mais de 70cm?

$$P(70 \le x \le 80) = (80-70)/(80-50) = 10/30 = 33,33\%$$

- Para uma Distribuição Uniforme com extremos a e b
 - Média

$$\mu = \frac{(a+b)}{2}$$

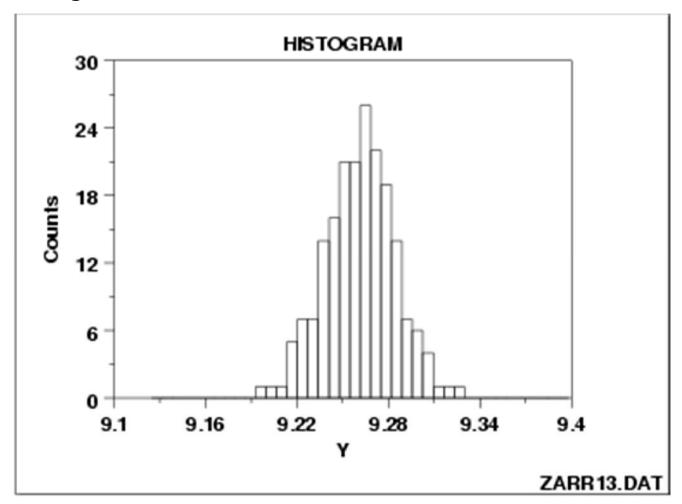
Variância

$$\sigma^2 = \frac{(b-a)^2}{12}$$

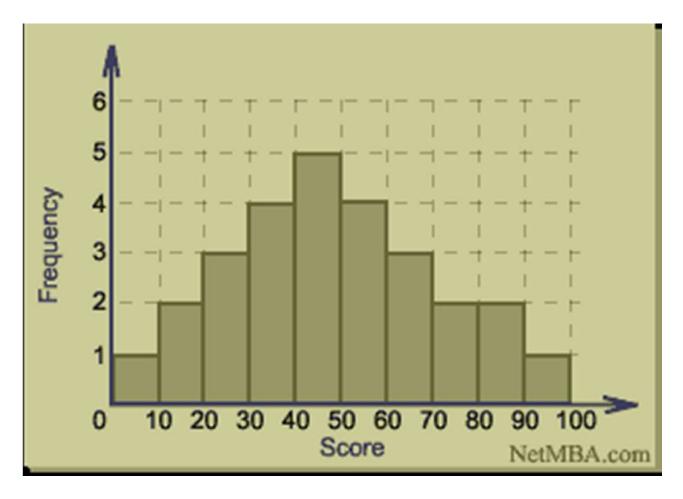
- Mais importante
- Com freqüência representam com boa aproximação fenômenos
 - Distribuições de freqüência de fenômenos naturais
- Probabilidades intervalares ao invés de pontuais
 - Probabilidade individual aproximadamente zero
- Servem como aproximação de probabilidades binomiais com grande n
- Distribuição de médias e de proporções em grandes amostras tendem a ser distribuições normais
 - Relevante implicação na amostragem

- Astrônomos no século XVIII
 - Mensurações consecutivas de um mesmo elemento tendiam a variar
 - massa de um objeto
 - Distância entre astros
 - Terra à Lua
 - Grande número de observações expostas em uma distribuição de freqüência
 - Valores em torno da média mais frequentes
 - Valores extremos mais escassos
 - → Distribuição Normal de Erros = Distribuição Normal

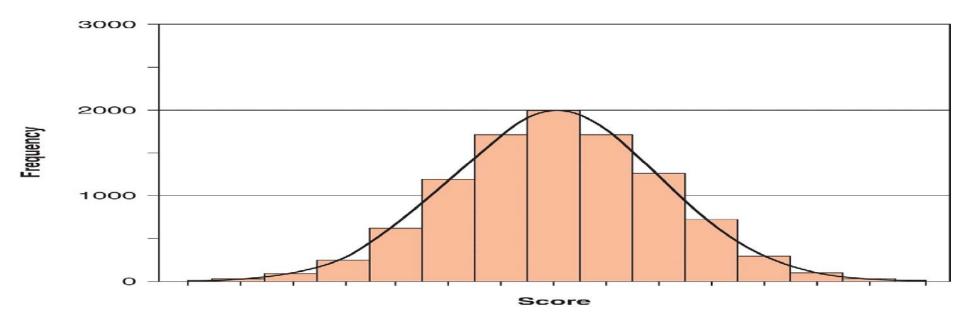
- Astrônomos no século XVIII
 - Grande número de observações expostas em uma distribuição de freqüência
 - Valores em torno da média mais frequentes
 - Valores extremos mais escassos
 - Comportamento similar para
 - Grandes amostras
 - Sub amostras
- Distribuição Gaussiana ou Normal
 - Karl F. <u>Gauss</u> (1777-1855)
 - Contribuição à teoria matemática



Distribuições de freqüência de observações freqüentemente acusavam a mesma forma



Distribuições de freqüência de observações freqüentemente acusavam a mesma forma



Uma curva contínua aproxima a distribuição de freqüências observadas

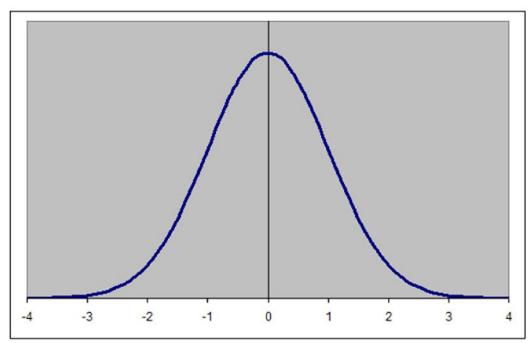
$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

- Características de uma Distribuição NORMAL
 - Quanto à Forma
 - Quanto a como é especificada
 - Quanto ao uso para obtenção de probabilidades

Características de uma Distribuição NORMAL

Quanto à Forma

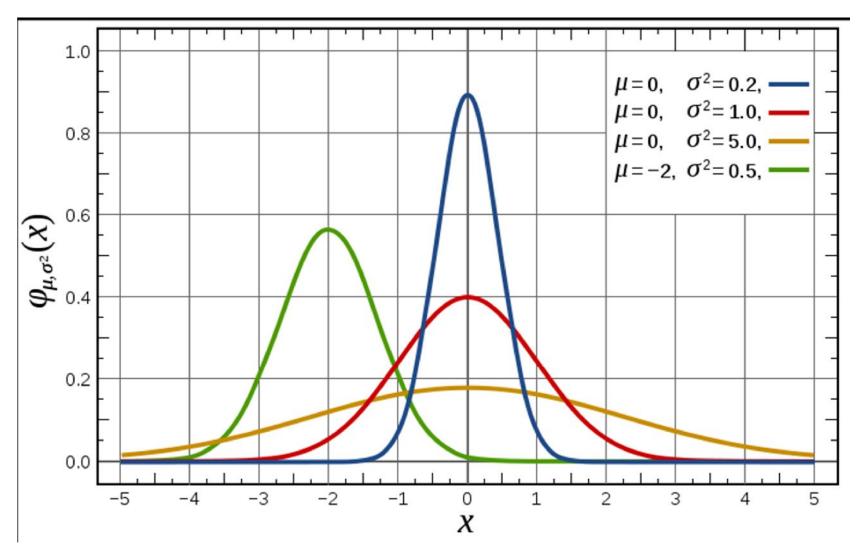
- Forma de sino
- Suave
- Unimodal
- Simétrica em relação à média
- Prolonga-se indefinidamente nas duas direções
- Tende a zero à medida que aumenta a distância a partir da média
- Teoricamente as observações podem variar de -∞ a +∞



- Características de uma Distribuição NORMAL
 - Quanto a como é especificada
 - Pode ser especificada pelos parâmetros
 - Média e desvio padrão
 - Assim, há uma única distribuição normal para cada combinação de média e desvio padrão
 - Médias e desvios padrões têm escalas contínuas =>
 - Ilimitadas distribuições normais

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

Características de uma Distribuição NORMAL

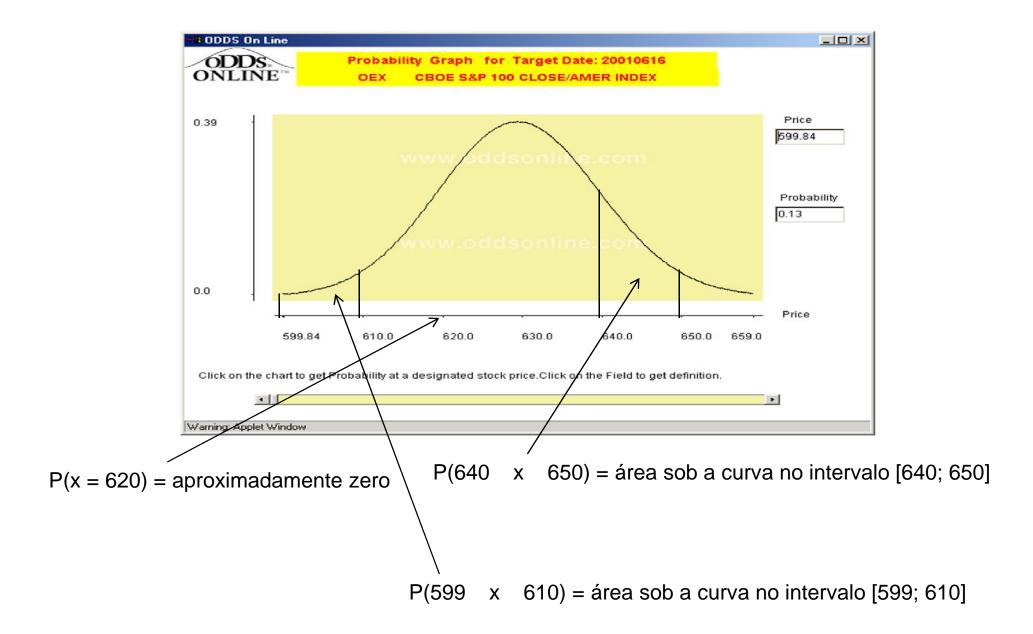


Distribuições Normais distintas associadas a combinações de média (µ) e desvio padrão/variância (²)

- Características de uma Distribuição NORMAL
 - Quanto ao uso para obtenção de probabilidades
 - Área sob a curva normal indica/representa probabilidades
 - Área total representa 100%
 - Probabilidade de obter qualquer valor da variável aleatória
 - Curva simétrica em relação à média
 - 50% de probabilidade de observar-se
 - Um valor superior à média
 - Um valor inferior à média
 - Probabilidade de um valor exato é aproximadamente zero
 - Escala de mensuração contínua (-∞ a +∞)
 - Probabilidade de um valor dentro de um intervalo
 - É a <u>área sob a curva normal compreendida entre os</u> <u>limites do intervalo</u>

- Distribuição NORMAL como Modelo
 - DN é uma distribuição teórica
 - Nenhum conjunto de valores ajusta-se 100%
 - Ideal para mensurações físicas numa distribuição de freqüência
 - Nem sempre valores reais variam de -∞ a +∞
 - Nem todos valores estão representados
 - Falha na mensuração ou coleta de dados
 - Facilidade de uso da DN para obter probabilidades compensa deficiências
 - DN é uma boa aproximação, um bom modelo, para dados reais

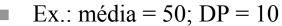
- DN é uma família infinita de distribuições
- Cada combinação de média e desvio padrão está associada a uma DN
- Não há tabelas para todas
- Alternativa é padronizar a DN
- Assim como o círculo com ponteiro tem área 100%
- A forma do círculo é importante, sendo o tamanho secundário
- A forma da DN é o importante
- A área sob a curva representa vale 1 (100%)

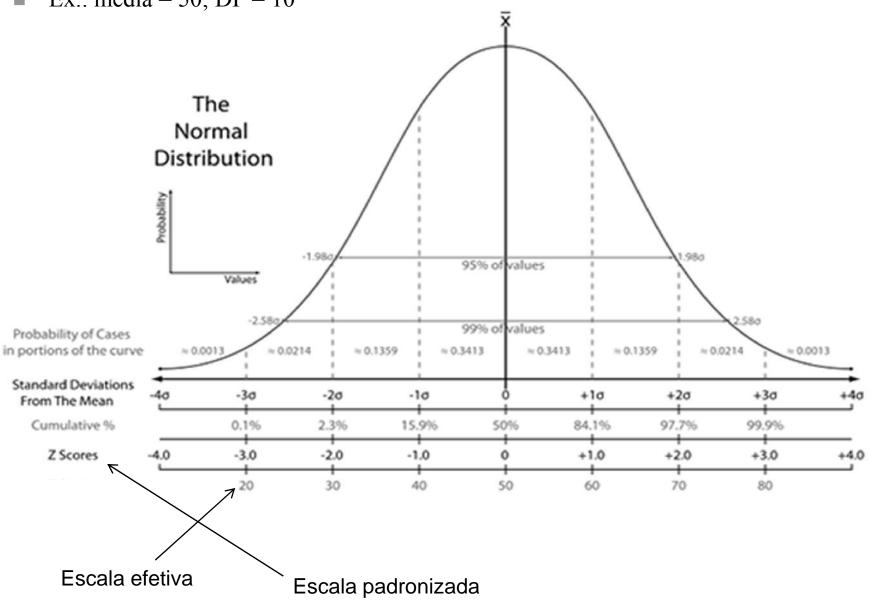


- A Distribuição NORMAL
 - É uma distribuição teórica
 - Populações e amostras não se comportam 100% como uma distribuição Normal
 - A distribuição Normal é uma aproximação
 - Uso da DN permite obter-se probabilidades
 - Ao dizer-se
 - Uma Variável Aleatória tem DN
 - Significa que a distribuição de frequências dos resultados possíveis pode ser satisfatoriamente aproximada por uma DN de probabilidades
 - Deste modo, a DN é um modelo

- Distribuição NORMAL PADRONIZADA
 - Esta característica da DN é crucial
 - Trabalha-se com valores relativos ao invés de valores reais de observações
 - Toma-se a média como referência (origem)
 - Desvio padrão como medida de afastamento da média
 - Unidade de medida é o desvio padrão
 - Nova escala z
 - Cada valor de observação é convertido em uma distância da média medida em desvio padrão (escore z)





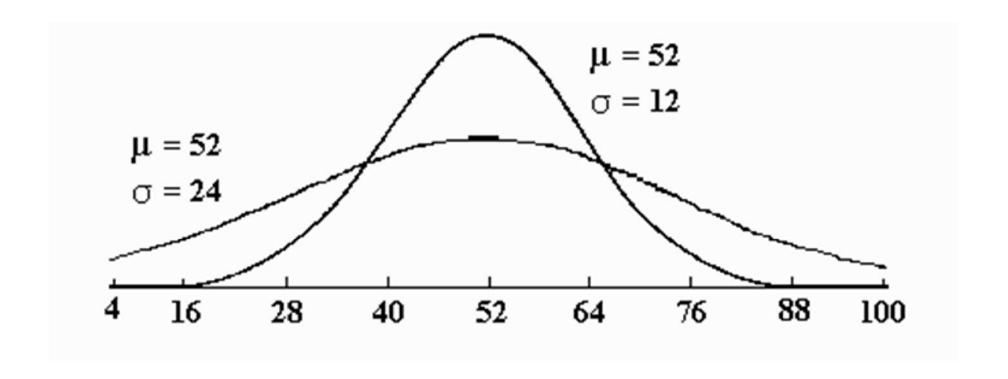


- Considere uma amostra com média 50 e desvio padrão 10
- 60 está
 - (60 50) = +10 unidades "acima" da média, ou
 - 10/10 = +1 desvio padrão acima da média
- 30 está
 - (30 50) = -20 unidades "abaixo" da média, ou
 - -20/10 = -2 DP "abaixo" da média
- 15 está
 - (15 50) = -35 unidades "abaixo" da média, ou
 - -35/10 = **-3,5 DP** "abaixo" da média

- Metodologia de padronização
- Converter diferença entre média e valor observado para um valor expresso em DP que indica a "distância" da observação à média

$$z = \frac{x - \mu}{\sigma}; \quad x = z\sigma + \mu$$

- Onde
 - **z** = número de DP a contar da média
 - $\mathbf{x} = \text{valor observado}$
 - µ = média
 - σ = desvio padrão

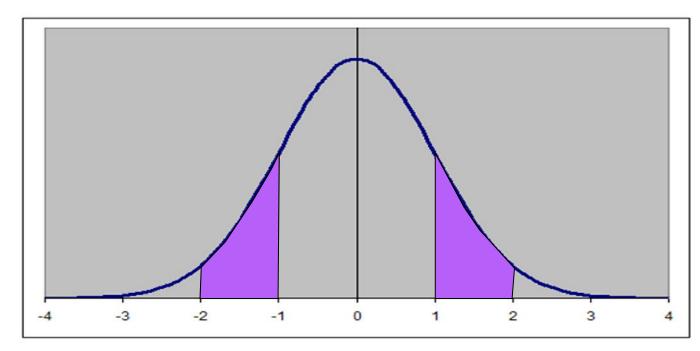


Duas Distribuições Normais com mesma média e DP distintos

- DN padronizada
 - Trabalha-se com valores relativos
 - Uma DN para qualquer distribuição que seja normal
 - Os valores de qualquer DN podem ser convertidos em escores z
 - Pode-se determinar as probabilidades de observações uma TABELA PADRONIZADA
 - Probabilidade é a área sob a curva

http://davidmlane.com/hyperstat/z table.html

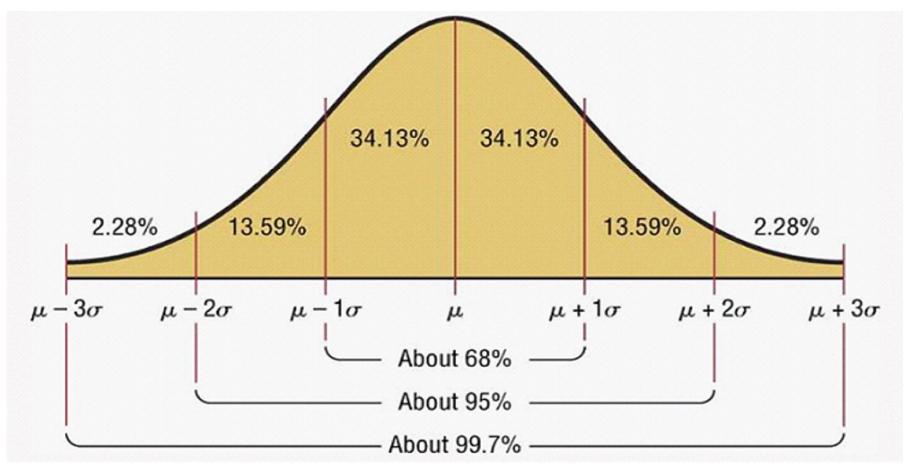
- DN simétrica em torno da média
 - Área total = 100%
 - Área à direita da média = 50%
 - Área à esquerda da média = 50%
 - Área entre -z e μ = Área entre μ e +z
 - Área entre -2z e -1z = Área entre +1z e +2z



- Tabela de Probabilidades (áreas) de uma DN Padronizada
- Tabela exibe área sob a curva (probabilidade)
 - Probabilidade de um valor estar naquele intervalo
- Tabelas
 - Área total
 - Área parcial (mais comum): 50%
- Exemplo área entre média e +z
 - Probabilidade de estar entre a média e determinado valor de z
 - P(0 < z < +2) = 0,4772 = 47,72%
 - P(0 < z < 1.83) = 0.4664
 - P(0 < z < 0) = 0,0000
 - P(0 < z < 3,14) = 0,4992

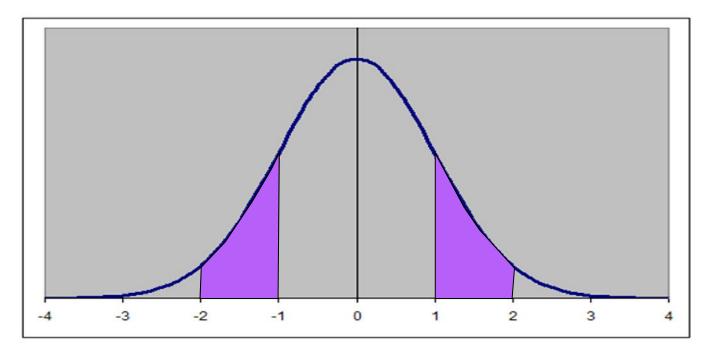
- De fato a DN é uma gama de distribuições
 - Combinações de médias e desvios padrões
 - Impossível elaborar-se tabelas de probabilidades para todas as DN
 - Experimentos mostraram que
 - Para uma Var. Aleatória com DN
 - 68% dos valores estão em $[-1\sigma; +1\sigma]$
 - 95% dos valores estão em [-2σ; +2σ]
 - 99,7% dos valores estão em $[-3\sigma; +3\sigma]$
 - Válido para qualquer DN

- Características de uma Distribuição NORMAL
 - Pelo fato de uma curva Normal poder ser especificada pela média e desvio padrão
 - Área entre um ponto qualquer e a média
 - É função do número de desvios padrões entre a média e o ponto



Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
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
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
Métodos	Quantitativos	0,4999 S	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999

- DN simétrica em torno da média
 - Área total = 100%
 - Área à direita da média = 50%
 - Área à esquerda da média = 50%
 - Área entre -z e μ = Área entre μ e +z
 - Área entre -2z e -1z = Área entre +1z e +2z



$$P(z > 0.01) = 0.5 - P(0 < z < 0.01) = 0.5 - 0.004 = 0.496 = 49.6\%$$

$$P(z > 0.53) = 0.5 - P(0 < z < 0.53) = 0.5 - 0.2019 = 0.2981 = 29.81\%$$

$$P(z > 2,42) = 0.5 - P(0 < z < 2,42) = 0.5 - 0.4922 = 0.0078 = 0.78\%$$

$$P(2 < z < 3) = P(0 < z < 3) - P(0 < z < 2) = 0.4987 - 0.4772 = 0.0215 = 2.15\%$$

$$P(1,42 < z < 2,41) = P(0 < z < 2,41) - P(0 < z < 1,42) = 0,492 - 0,4222 = 0,0698$$

$$P(3 < z < 3.5) = P(0 < z < 3.5) - P(0 < z < 3) = 0.4998 - 0.4987 = 0.0011 = 0.11\%$$

- Lembrando a simetria da Distribuição Gaussiana e que estamos olhando uma tabela que representa 50%
- P(-1 < z < 0) = P(0 < z < +1) = 0.3413
- P(-2,15 < z < 0) = P(0 < z < +2,15) = 0,4842
- P(-3,16 < z < 0) = P(0 < z < +3,16) = 0,4992
- P(z < -2,42) = 0.5 P(0 < z < 2,42) = 0.5 0.4922 = 0.0078 = 0.78%
- P(z < -3,16) = 0.5 P(0 < z < +3.16) = 0.5 0.4992 = 0.0008 = 0.08%
- P(-1 < z < 1) = P(-1 < z < 0) + P(0 < z < 1) = P(0 < z < +1) + P(0 < z < 1) = 0,3413 + 0,3413 = 0,6826
- P(-1,92 < z < 1,95) = P(-1,92 < z < 0) + P(0 < z < 1,95) = P(0 < z < 1,92) + P(0 < z < 1,95) = 0,4726 + 0,4744 = 0,947 = 94,7%
- P(-2,42 < z < 2,97) = P(-2,42 < z < 0) + P(0 < z < 2,97) = 0,4922 + 0,4985 = 0,9907 = 99,07%

- Para determinada população com Distribuição Normal
 - Probabilidade de valor efetivo x é calculado
 - Transformando-se x em z

$$z = \frac{x - \mu}{\sigma}; \quad x = z\sigma + \mu$$

- Para determinada população com Distribuição Normal
 - Média = 80 e Desvio Padrão = 10

$$x = 80 \rightarrow z = (80 - 80) / 10 \rightarrow z = 0$$

$$x = 90 \rightarrow z = (90 - 80) / 10 \rightarrow z = 1$$

$$x = 70 \rightarrow z = (70 - 80) / 10 \rightarrow z = -1$$

$$x = 100 \rightarrow z = (100 - 80) / 10 \rightarrow z = 2$$

$$x = 60 \rightarrow z = (60 - 80) / 10 \rightarrow z = -2$$

$$x = 110 \Rightarrow z = (110 - 80) / 10 \Rightarrow z = 3$$

$$x = 50 \rightarrow z = (50 - 80) / 10 \rightarrow z = -3$$

Para determinada população com Distribuição Normal

Média = 80 e Desvio Padrão = 10

$$x = 80 = > z = 0$$

 $x = 90 = > z = 1$
 $P(80 < x < 90) = P(0 < z < 1) = 0,3413$
 $x = 100 = > z = 2$
 $P(80 < x < 100) = P(0 < z < 2) = 0,4772$
 $x = 110 = > z = 3$
 $P(80 < x < 110) = P(0 < z < 3) = 0,4986$
 $x = 70 = > z = -1$
 $P(70 < x < 80) = P(-1 < z < 0) = 0,3413$
 $x = 60 = > z = -2$
 $P(60 < x < 80) = P(-2 < z < 0) = 0,4772$
 $x = 50 = > z = -3$
 $P(50 < x < 80) = P(-3 < z < 0) = 0,4986$

Para determinada população com Distribuição Normal

```
Média = 80 e Desvio Padrão = 10
  x = 70 ==> z = -1
  x = 90 ==> 7 = 1
P(70 < x < 90) = P(-1 < z < 1) = P(-1 < z < 0) +
P(0 < z < 1) = 0.3413 + 0.3413 = 0.6826
  x = 60 ==> z = -2
  x = 100 ==> 7 = 2
P(60 < x < 100) = P(-2 < z < 2) = P(-2 < z < 0) +
P(0 < z < 2) = 0.4772 + 0.4772 = 0.9544
  x = 50 = > 7 = -3
  x = 110 =  7 = 3
P(50 < x < 110) = P(-3 < z < 3) = P(-3 < z < 0) +
P(0 < z < 3) = 0.4986 + 0.4986 = 0.9972
```

Para determinada população com Distribuição Normal

Para determinada população com Distribuição Normal

Padronização de uma variável x

observação	Z	observação	Z	
32	-0,09925	15	-0,74543	
32	-0,09925	16	-0,70742	
31	-0,13726	16	-0,70742	
17	-0,66941	17	-0,66941	
21	-0,51737	17	-0,66941	
46	0,432899	17	-0,66941	
93	2,219398	18	-0,6314	
18	-0,6314	21	-0,51737	
16	-0,70742	23	-0,44135	
15	-0,74543	26	-0,32731	
101	2,523483	31	-0,13726	
16	-0,70742	32	-0,09925	
17	-0,66941	32	-0,09925	
17	-0,66941	34	-0,02323	
26	-0,32731	46	0,432899	
23	-0,44135	68	1,269132	
68	1,269132	93	2,219398	
34	-0,02323	101	2,523483	

Média: 34,6111111

DP: 26,3084395

Distribuição EXPONENCIAL

- Distribuição Exponencial
 - Probabilidades em intervalo contínuo
 - De tempo
 - De distância
 - Número de eventos no intervalo especificado
 - Permite aproximar a probabilidade de N ocorrências no intervalo dado
 - Falhas de um equipamento eletrônico
 - Tempo entre chegada de clientes
 - Tempo de atendimento a cliente
 - Chamadas de emergência em determinado dia da semana
 - Muito próxima à Distribuição de Poisson

Distribuição EXPONENCIAL

- Distribuição Exponencial
 - Probabilidades em intervalo contínuo
 - De tempo
 - De distância
 - Muito próxima à Distribuição de Poisson

