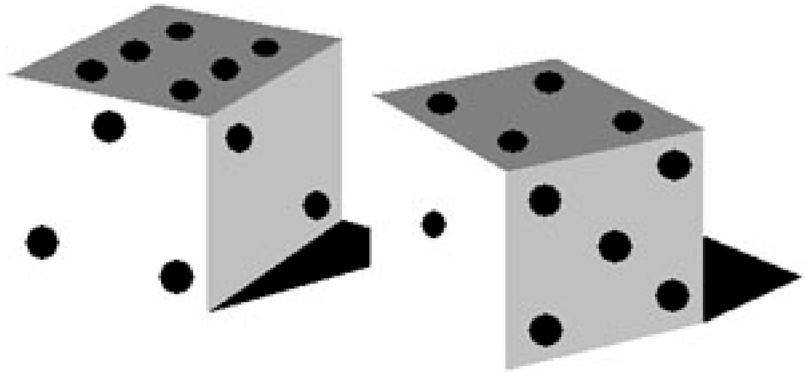
Probabilidade



Prof. Lorí Viali, Dr.

viali@mat.pucrs.br

http://www.pucrs.br/~viali/

4

Porto Alegre, agosto de 2002

Variável Aleatória Contínua

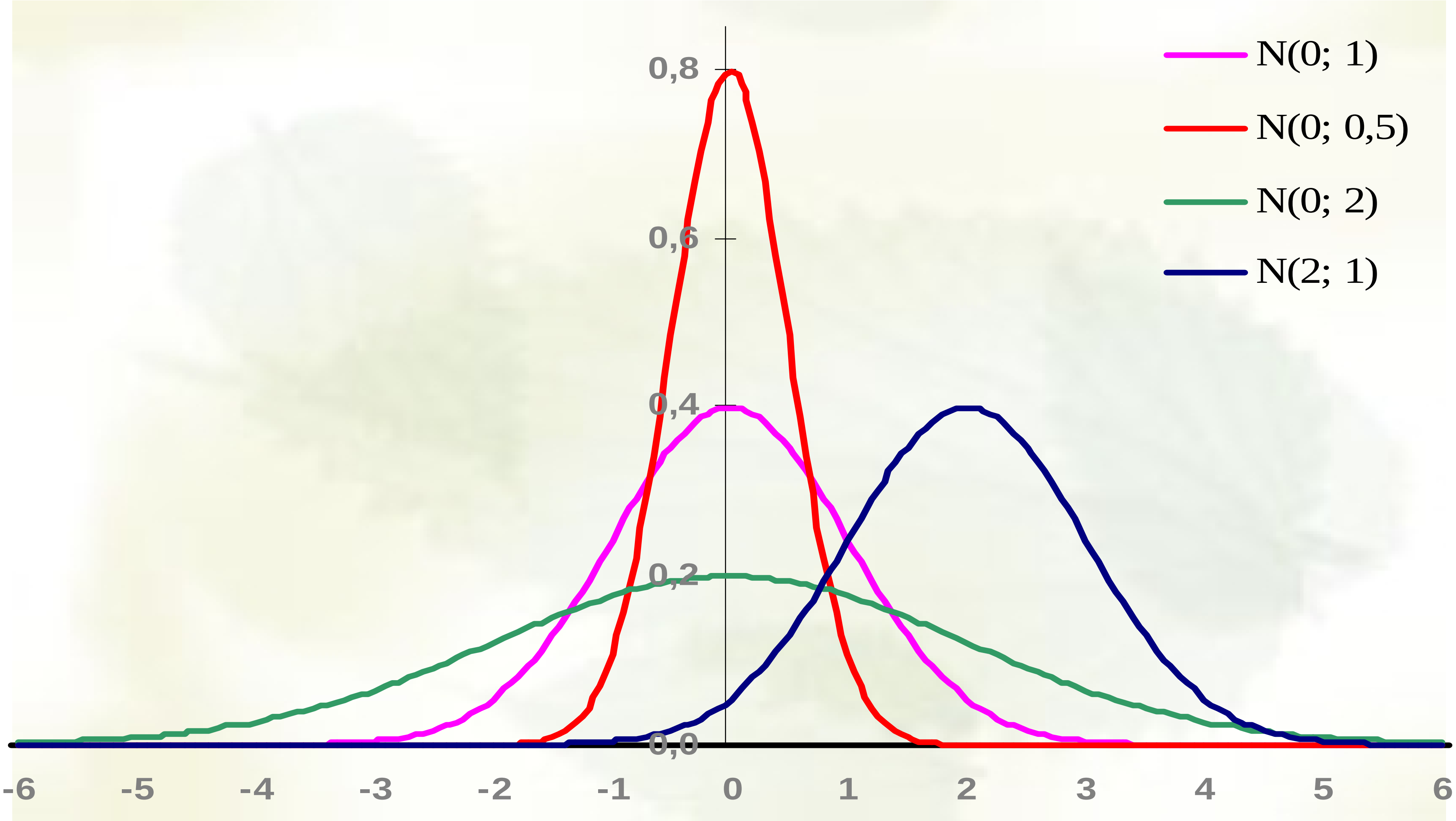
A Distribuição Normal

Uma variável aleatória X tem uma distribuição **normal** se sua fdp for do tipo:

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

com $-\infty < \mu < \infty$ e $\sigma > 0$

Gráficos



Cálculo da Probabilidade

$$P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}.\sigma}.e^{-\frac{1}{2}.\left(\frac{u-\mu}{\sigma}\right)^2} du = ?$$

A normal não é integrável através do TFC, isto é, não existe F(x) tal que F'(x) = f(x).

Solução do Problema

Utilizar integração numérica. Como não é possível fazer isto com todas as curvas, escolheu-se uma para ser tabelada (integrada numericamente).

A normal padrão

A curva escolhida é a $N(0, 1)$, isto é, com $\mu = 0$ e $\sigma = 1$.

Se X é uma $N(\mu, \sigma)$, então:

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

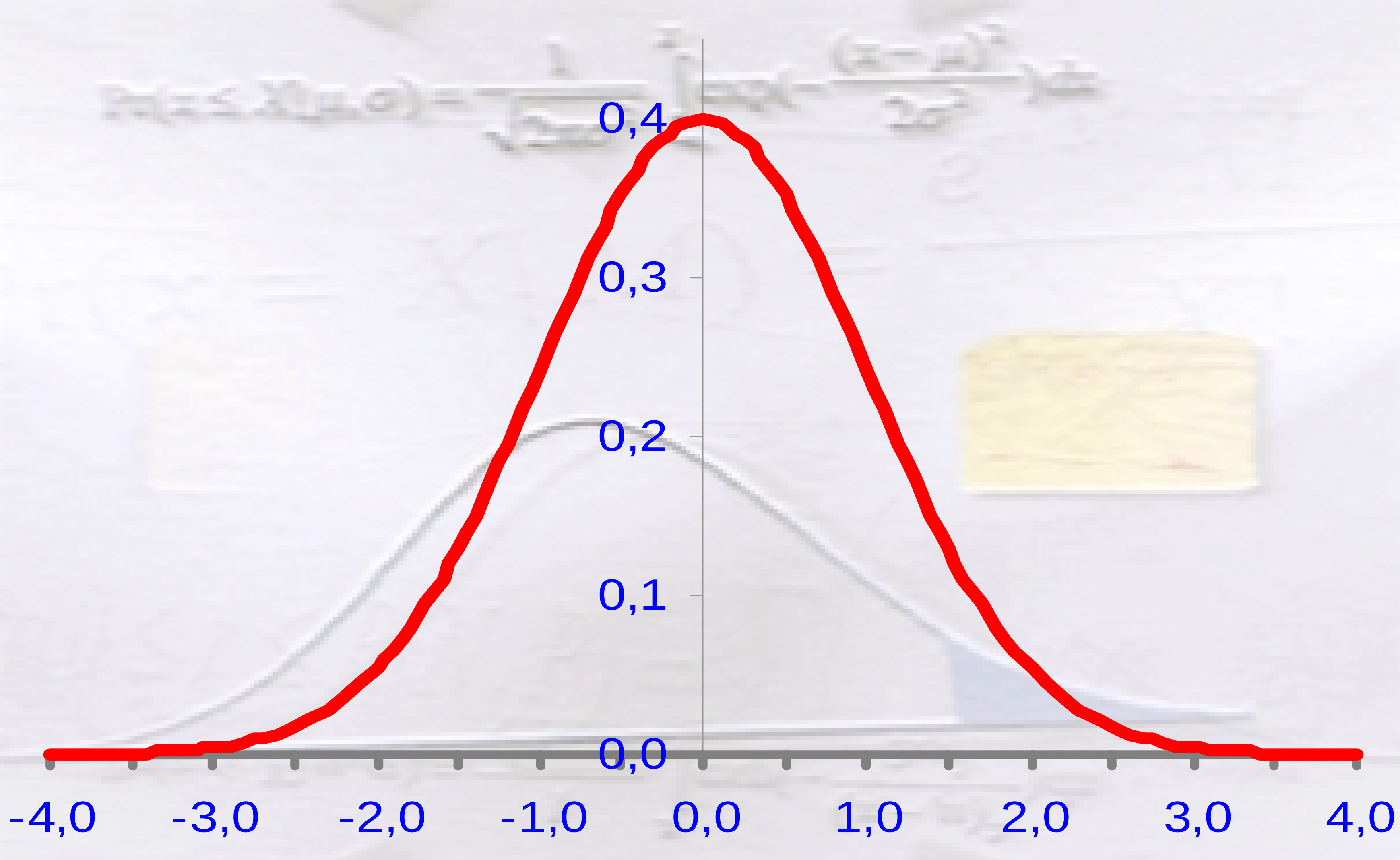
Será uma $N(0; 1)$

A fdp da variável Z é dada por:

$$\varphi(z) = \frac{1}{\sqrt{2\pi}}.e^{-\frac{z^2}{2}.},\ z \in \Re$$

uma vez que $\mu = 0$ e $\sigma = 1$.

Distribuição N(0, 1)



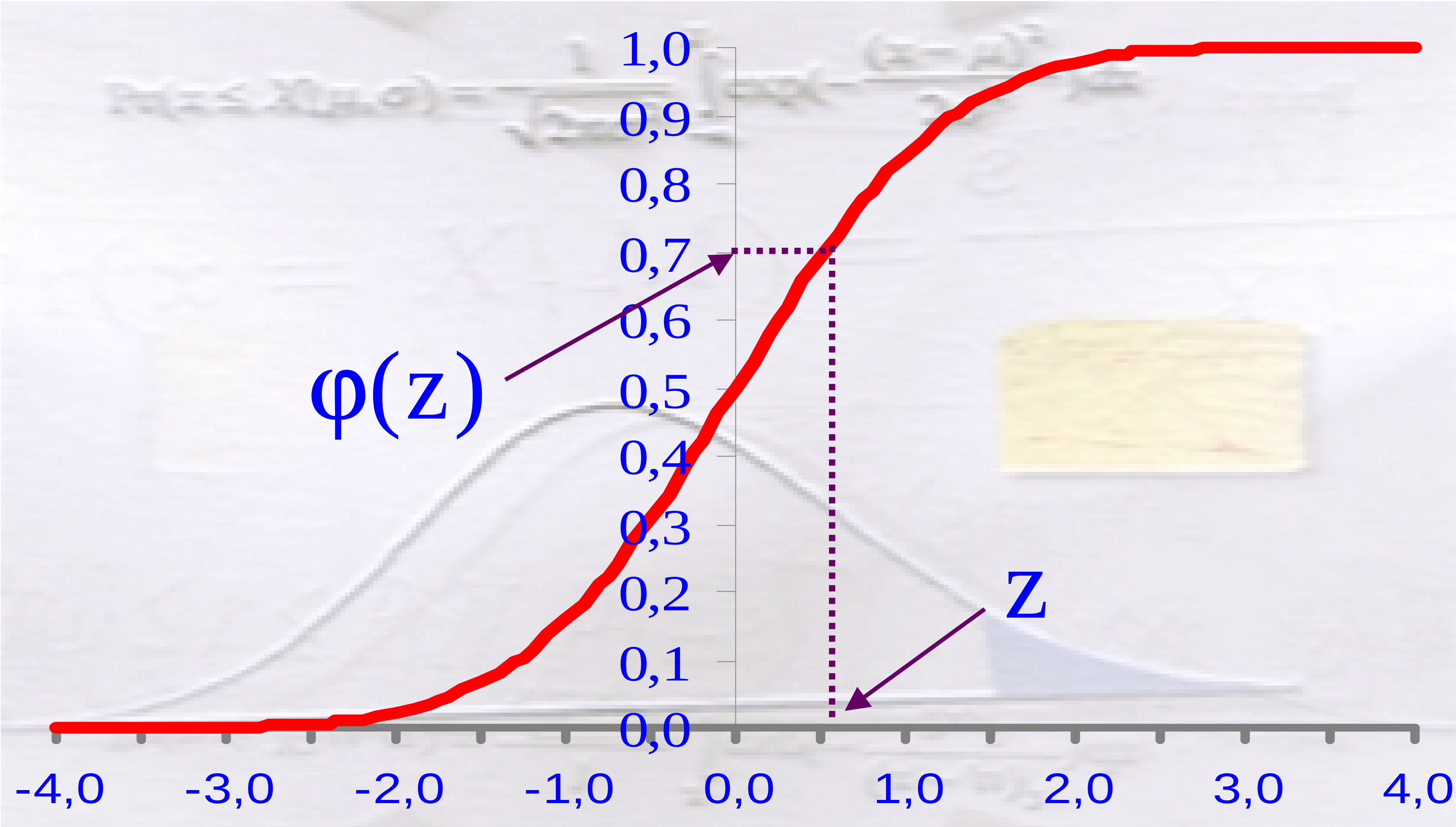
Tabela

O que é tabelado é a FDA da variável Z, isto é:

$$P(Z \le z) = \int_{\infty}^{z} \varphi(u)du =$$

$$= \int_{\infty}^{z} \frac{1}{\sqrt{2\pi}}.e^{-\frac{u^2}{2}}.du = \Phi(z)$$

FDA da N(0; 1)



Utilização da Tabela

Área à esquerda (abaixo) de "z"

$$P(Z \leq z) = \Phi(z) = \text{Leitura direta}$$

Área à direita (acima) de "z"

$$P(Z > z) = 1 - P(Z \leq z) = 1 - \Phi(z) = \Phi(-z)$$

Área entre dois valores de "z"

$$P(z_1 < Z < z_2) = \Phi(z_2) - \Phi(z_1)$$

A tabela é construída como uma matriz. As linhas fornecem a unidade ou unidade mais décimo e as colunas fornecem os centésimos.

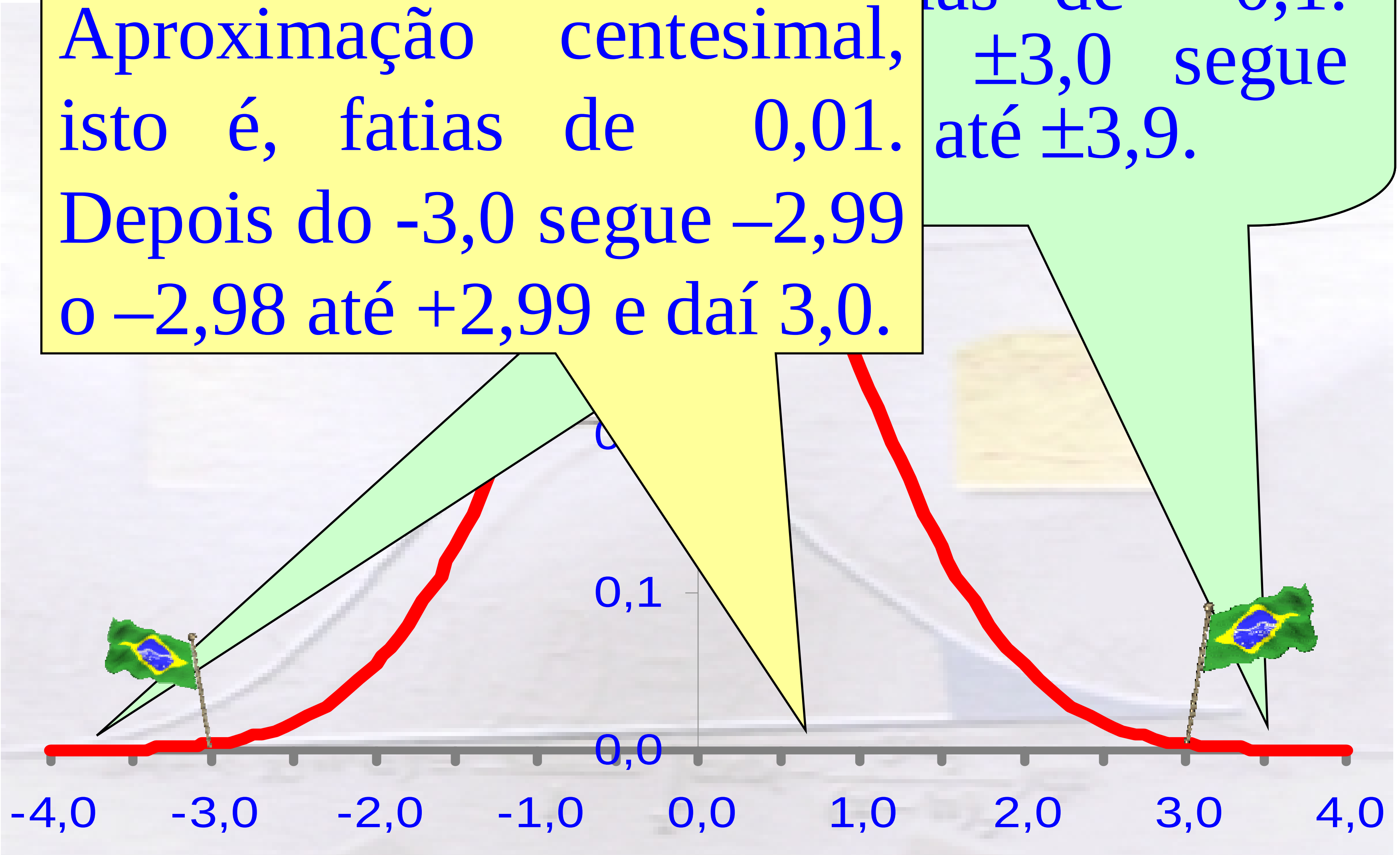
Assim para ler, por exemplo, -0,15 deve-se procurar na linha do –0,1 + coluna do 5 (sexta coluna). A primeira é a do “0” (zero).

A aproximação é centesimal (2 casas após a vírgula) exceto na linha do –3 e do +3, que estão destacadas, onde a aproximação é, em virtude da pouca área, **decimal.** Observe que está escrito –3 e não –3,0!

Tabela

Aproximação decimal, isto é, fatias de 0,1. ... ±3,0 segue ... até ±3,9.

Aproximação centesimal, isto é, fatias de 0,01. Depois do -3,0 segue –2,99 o –2,98 até +2,99 e daí 3,0.



z	0	1	2	3
-3	0,0013	[illegible]	0,0007	0,0005
-2,9	[illegible]	[illegible]	[illegible]	0,0017
-2,8	[illegible]	[illegible]	0,0024	0,0023
-2,7	0,0035	0,0034	0,0033	0,0032
-2,6	[illegible]	[illegible]	0,0044	0,0043
-2,5	[illegible]	[illegible]	[illegible]	0,0057
-2,4	0,0082	[illegible]	0,0078	0,0075
-2,3	0,0107	0,0104	[illegible]	0,0099
-2,2	0,0139	[illegible]	[illegible]	[illegible]
-2,1	0,0179	[illegible]	[illegible]	0,0166
-2,0	0,0228	0,0222	0,0217	0,0212

$P(Z < -3,3) = \Phi(-3,3)$

$P(Z < -2,53) = \Phi(-2,53)$

$P(Z < -2,00) = \Phi(-2,00)$

Exemplo

Uma VAC tem distribuição normal de média 50 e desvio padrão 8. Determinar:

(a) $P(X \leq 40)$

(b) $P(X > 65)$

(c) $P(45 < X < 62)$

(a) $P(X \leq 40)$

$$P(X \leq 40) = P\left(\frac{X-\mu}{\sigma} \leq \frac{40-50}{8}\right) =$$

$$= P(Z \leq -1{,}25) = 10{,}56\%$$

(b) $P(X > 65)$

$$P(X > 65) = P\left(\frac{X - \mu}{\sigma} > \frac{65 - 50}{8}\right) =$$

$$= P(Z > 1{,}88) = 1 - P(Z < 1{,}88) =$$

$$= 1 - \Phi(1{,}88) = \Phi(-1{,}88) = 3{,}01\%$$

(c) $P(45 < X < 62)$

$$P(45 < X < 62) =$$

$$= P\left(\frac{45-50}{8} < \frac{X-\mu}{\sigma} < \frac{62-50}{8}\right) =$$

$$= P(-0{,}62 < Z < 1{,}50) =$$

$$= \Phi(1{,}50) - \Phi(-0{,}62) =$$

$$= 93{,}32\% - 26{,}76\% = 66{,}56\%$$

A função Inversa

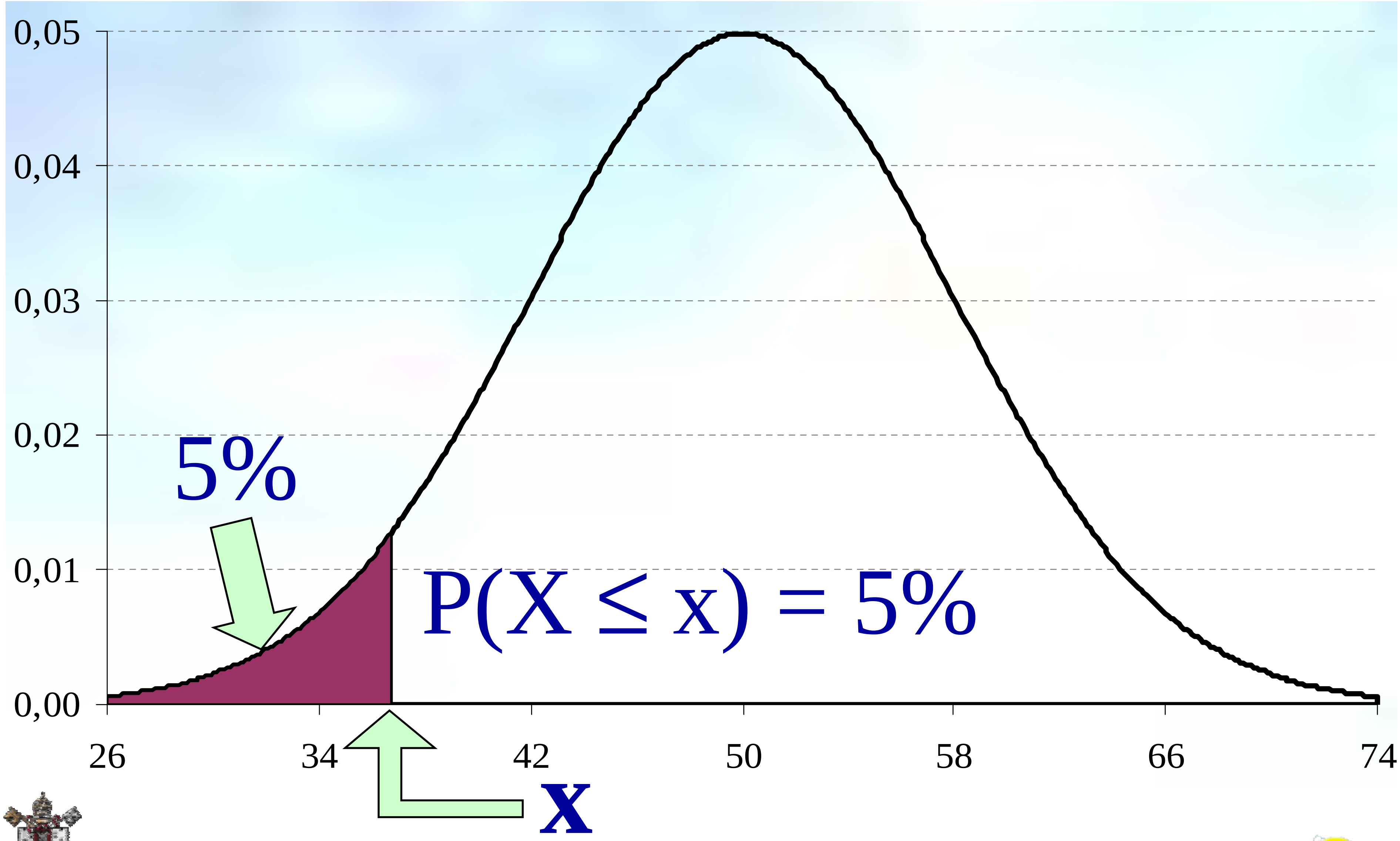
Uma VAC tem distribuição normal de média 50 e desvio padrão 8. Determinar:

(a) $P(X \leq x) = 5\%$

(b) $P(X > x) = 1\%$

Para resolver este tipo de exercício é preciso utilizar a função inversa, isto pode ser feito direto na tabela. Só que agora devemos procurar uma probabilidade (corpo da tabela) e obter um valor de “z” (lateral da tabela).

Graficamente, tem-se:



Em **(a)** temos $P(X \le x) = 5\%$

$$P(X \le x) = P\left(\frac{X - \mu}{\sigma} \le \frac{x - 50}{8}\right) =$$

$$= P(Z \le z) = \Phi(z) = 5\%$$

onde $\quad z = \dfrac{x - 50}{8}$

Se $\Phi(z) = 5\%$, então

$$\Phi^{-1}[\Phi(z)] = \Phi^{-1}(5\%)$$

$$z = \Phi^{-1}(0{,}05)$$

Procurando na tabela, o valor (z) mais próximo de 5% = 0,05, tem-se:

z	0	1	2	3	4	5
-3	0,0013	0,0010	0,0007	0,0005	0,0003	0,0002
-2,9	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016
-2,8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022
-2,7	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030
-2,6	0,0047	0,0045	0,0044	0,0043	0,0041	0,0040
-2,5	0,0062	0,0060	0,0059	0,0057	[illegible]	[illegible]
-2,4	0,0082	0,0080	[illegible]	[illegible]	[illegible]	[illegible]
-2,3	0,0107	0,0104	[illegible]	[illegible]	[illegible]	[illegible]
-2,2	0,0139	0,0136	0,0132	[illegible]	0,0125	0,0122
-2,1	0,0179	0,0174	0,0170	[illegible]	0,0162	0,0158
-2,0	0,0228	0,0222	0,0217	[illegible]	0,0207	0,0202
-1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256
-1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322
-1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401
-1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495
-1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606

z = -1,64

z = -1,65

Como os dois valores estão a mesma distância, isto é, apresentam o mesmo erro (0,0005), pega-se a média entre eles.

Assim

$$z = \frac{1{,}64 + 1{,}65}{2} = 1{,}645$$

Como $z = \frac{x - 50}{8}$, tem − se :

$$1{,}645 = z = \frac{x - 50}{8} \Rightarrow$$

$$x = 50 - 1{,}645 \,.\, 8 = 36{,}84$$

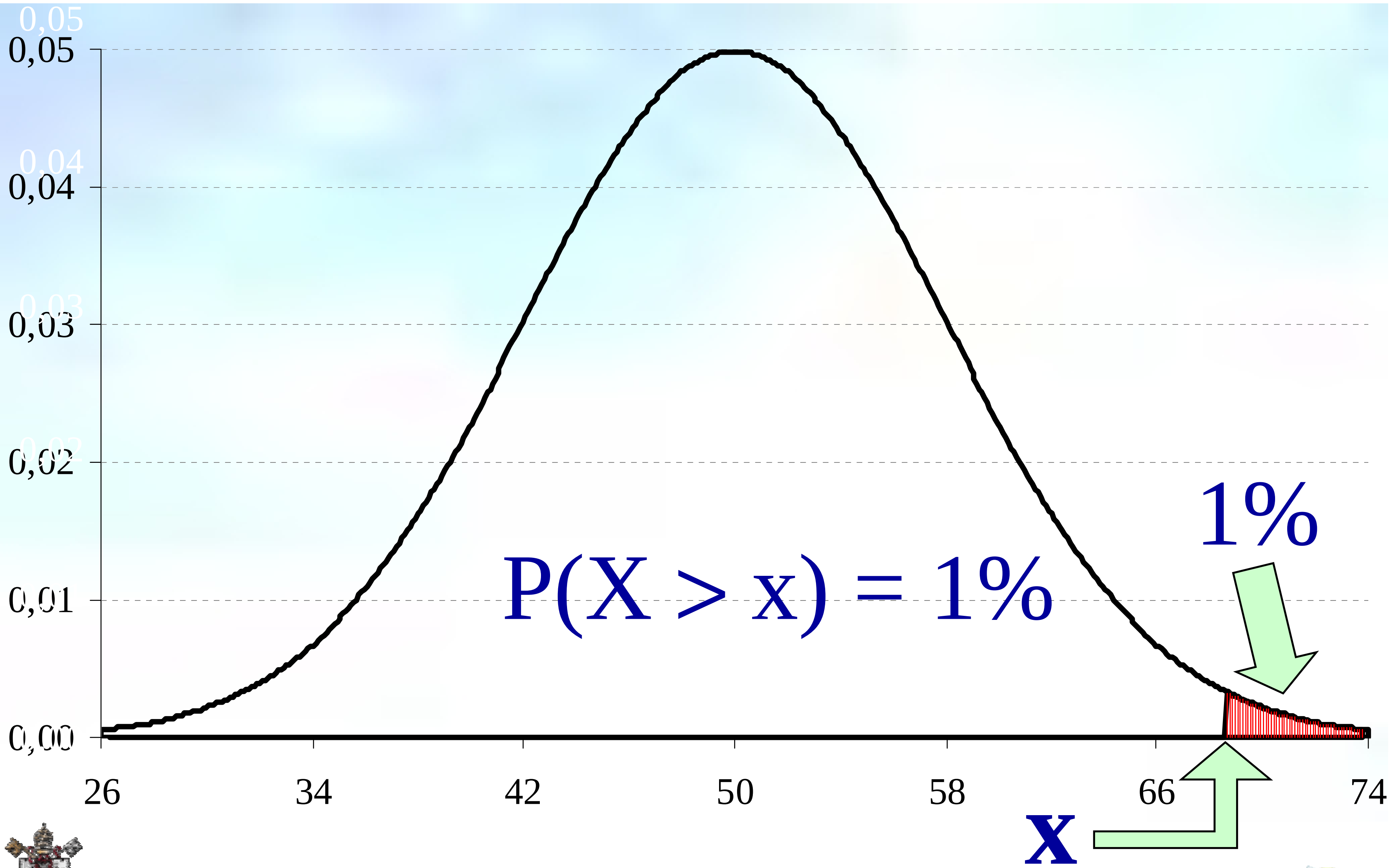
Em **(b)** temos P(X > x) = 1%

$$P(X > x) = P\left(\frac{X-\mu}{\sigma} > \frac{x-50}{8}\right) =$$

$$= P(Z > z) = 1 - \Phi(z) = 1\% = 0{,}01$$

$$\text{Mas} \quad 1 - \Phi(z) = \Phi(-z)$$

$$\text{Logo } -z = \Phi^{-1}(0{,}01)$$


0,05
0,04
0,03
0,02
0,01
0,00
26
34
42
50
58
66
74
P(X > x) = 1%
1%
x

Procurando na tabela, o valor (z) mais próximo de 1% = 0,01, tem-se:

z = -2,33

Conforme pode ser visto na próxima lâmina!

z	0	1	2	3
-3	0,0013	0,0010	0,0007	0,0005
-2,9	0,0019	0,0018	0,0018	0,0017
-2,8	0,0026	0,0025	0,0024	0,0023
-2,7	0,0035	[illegible]	[illegible]	0,0032
-2,6	0,0047	[illegible]	[illegible]	0,0043
-2,5	0,0062	0,0060	[illegible]	0,0057
-2,4	0,0082	0,0080	0,0078	0,0075
-2,3	0,0107	0,0104	0,0102	0,0099
-2,2	0,0139	0,0136	0,0132	0,0129
-2,1	0,0179	0,0174	0,0170	0,0166
-2,0	0,0228	0,0222	0,0217	0,0212

z = -2,33

Como

$$-z = \Phi^{-1}(0{,}01),\ \text{tem} - \text{se}:$$

$$-(-2{,}33) = \frac{x - 50}{8} \Rightarrow$$

$$x = 2{,}33.8 + 50 = 68{,}64$$

Outras Distribuições

Student ou t de Student

Uma variável aleatória X tem uma distribuição **“t”** ou de **Student** se sua fdp for do tipo:

$$f(x) = \frac{\Gamma\left(\frac{\upsilon + 1}{2}\right)\left(1 + \frac{x^2}{\upsilon}\right)^{-\frac{\upsilon + 1}{2}}}{\sqrt{\pi\upsilon}\,.\Gamma\left(\frac{\upsilon}{2}\right)}$$

para $x \in \Re$

onde Γ é a função dada por

$$\Gamma(p) = \int_0^{\infty} x^{p-1} e^{-x}\, dx$$

para $p > 0$

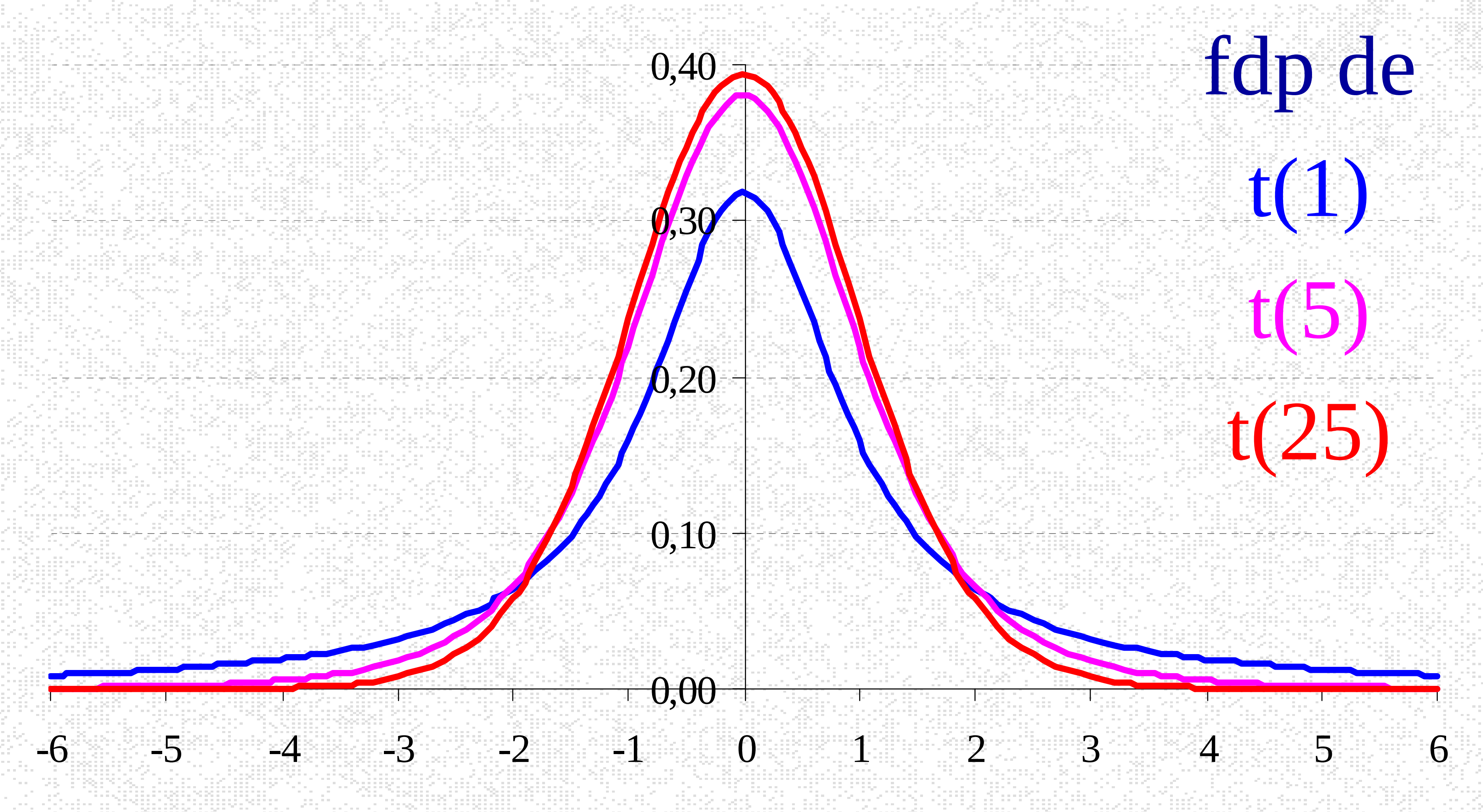
$$\Gamma(p) = (p - 1)\, \Gamma(p - 1)$$

$$\Gamma(1/2) = \sqrt{\pi}$$

$$\Gamma(n) = (n - 1)!$$

se $n \in Z$

Gráficos



Caracterização

Expectância ou Valor esperado

$$\mu = E(X) = 0$$

Variância

$$Var(X) = \frac{\upsilon}{\upsilon - 2}$$

O valor υ é denominado de "Grau de liberdade"

Tabelas

O que é tabelado é a função inversa (**percentis**), em relação a área à direita (**unilateral**) de cada curva (**uma para cada linha**), ou a soma das caudas (**bilateral**), isto é, a tabela retorna um valor "**t**" tal que $\mathbf{P(T \geq t) = \alpha}$ (**unilateral**) ou $\mathbf{P(|T| \geq t) = \alpha}$.

As duas opções podem ser colocadas em uma mesma tabela. Pode-se ler uma área (α) de cima para baixo e se ter um valor unilateral (**P(T ≥ t**) = α) ou ler a área (α) de baixo para cima e se ter um valor "t" tal que **P(T ≥ t**) = $\alpha/2$.

	0,200	0,100	0,050	0,040	0,030	0,020
1	3,078	6,314	12,706	15,894	21,205	31,821
2	1,886	2,920	4,303	4,849	5,643	6,965
3	1,638	[illegible]	[illegible]	[illegible]	3,896	4,541
4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	3,747
5	[illegible]	[illegible]	[illegible]	[illegible]	3,003	3,365
6	1,440	[illegible]	[illegible]	2,612	2,829	3,143
7	1,415	1,895	[illegible]	2,517	2,715	2,998
8	1,397	1,860	[illegible]	2,449	2,634	2,896
9	1,383	1,833	2,262	2,398	2,574	2,821
10	1,372	1,812	2,228	2,359	2,527	2,764

$P(|T_9| \geq 2,262) = 5\%$

	0,200	0,100	0,050	0,040	0,030	0,020
1	3,078	6,314	12,706	15,894	21,205	31,821
2	1,886	2,920	4,303	4,849	5,643	6,965
3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	4,541
4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	3,747
5	[illegible]	[illegible]	[illegible]	[illegible]	3,003	3,365
6	1,440	1,943	[illegible]	2,612	2,829	3,143
7	1,415	1,895	[illegible]	2,517	2,715	2,998
8	1,397	1,860	[illegible]	2,449	2,634	2,896
9	1,383	1,833	2,262	2,398	2,574	2,821
10	1,372	1,812	2,228	2,359	2,527	2,764

$P(T_9 < -2,262) = 2,5\%$

ou

$P(T_9 > 2,262) = 2,5\%$

Qui-Quadrado

Uma variável aleatória X tem uma distribuição **Qui-Quadrado** se sua fdp for do tipo:

$$f(x) = \begin{cases} \dfrac{x^{\frac{\upsilon}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{\upsilon}{2}} \Gamma\left(\dfrac{\upsilon}{2}\right)} & \text{se } x > 0 \\ 0 & \text{se } x \leq 0 \end{cases}$$

Caracterização

Expectância ou Valor esperado

$$E(X) = \upsilon$$

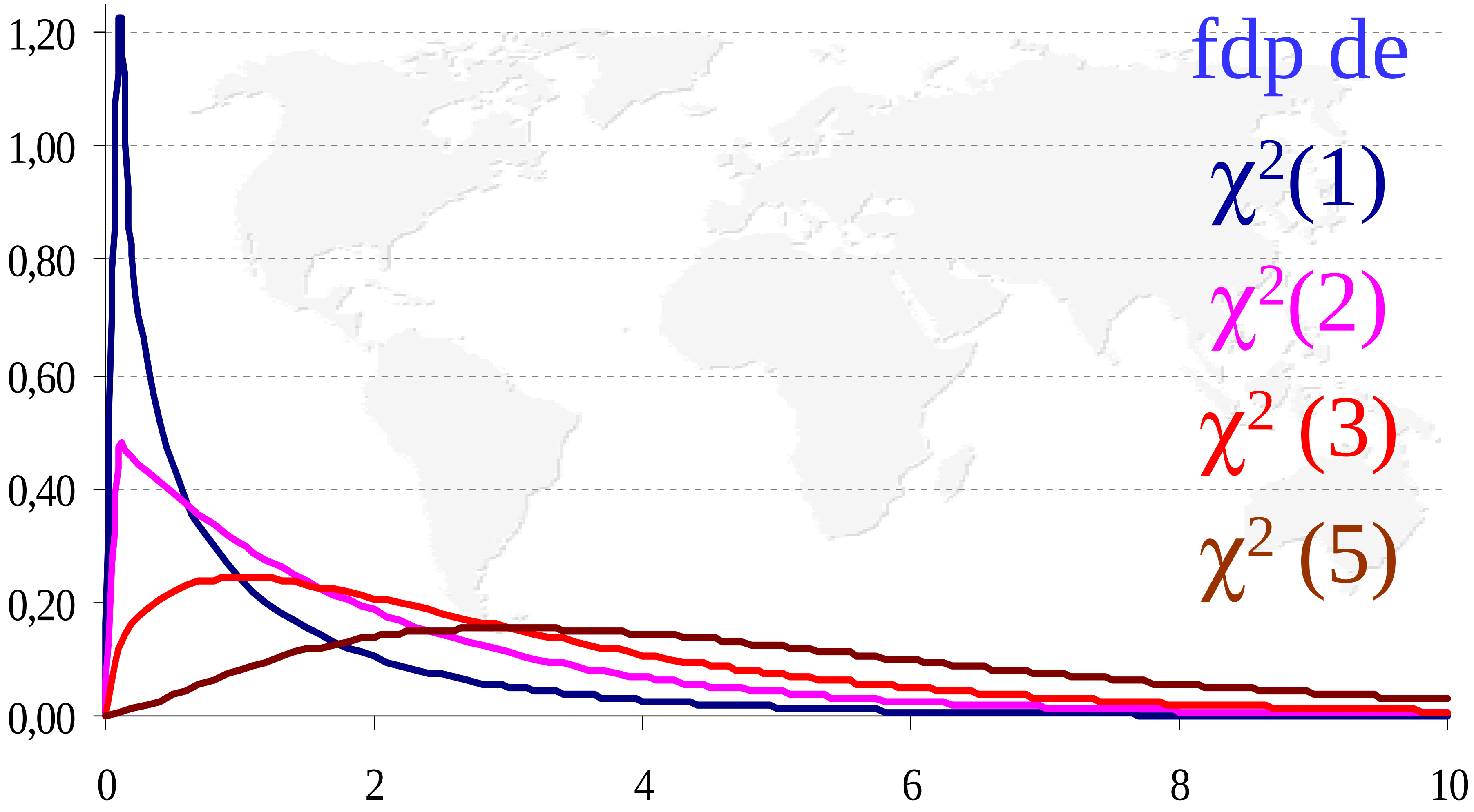
Variância

$$Var(X) \quad = 2\upsilon$$

O valor υ é denominado de

"Grau de liberdade"

Gráficos



Tabelas

O que é tabelado é a função inversa, em relação a área à direita de cada curva (**uma para cada linha**), isto é, dado um valor de área na cauda direita (α), a tabela retorna um valor "**x**" tal que $\mathbf{P(\chi^2 \geq x) = \alpha}$

	0,995	**0,990**	**0,975**	**0,950**	**0,900**
1	0,000	0,000	0,001	0,004	0,016
2	0,010	0,020	0,051	0,103	0,211
3	0,072	0,115	0,216	0,352	0,584
4	0,207	0,297	0,484	[illegible]	1,064
5	0,412	0,554	[illegible]	[illegible]	1,610
6	[illegible]	[illegible]	[illegible]	[illegible]	2,204
7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
8	1,344	[illegible]	[illegible]	[illegible]	3,490
9	1,735	2,088	2,700	3,325	4,168
10	2,156	2,558	3,247	3,940	4,865

$P[\chi^2(2) \geq 0{,}211] = 90\%$

	0,100	0,050	0,025	0,010	0,005
41	52,949	56,942	60,561	64,950	68,053
42	[illegible]	[illegible]	[illegible]	[illegible]	69,336
43	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
44	56,369	[illegible]	[illegible]	68,710	71,892
45	57,505	61,656	[illegible]	69,957	73,166
46	58,641	62,830	[illegible]	71,201	74,437
47	59,774	64,001	67,821	72,443	75,704
48	60,907	65,171	69,023	[illegible]	76,969
49	62,038	66,339	70,222	74,919	78,231
50	63,167	67,505	71,420	76,154	79,490

$P[\chi^2(49) \geq 74,919] = 1\%$

Snedecor ou F

Uma variável aleatória X tem uma distribuição **“F”** ou de **Snedecor** se sua fdp for do tipo:

$$f(x) = \begin{cases} \dfrac{\Gamma\left(\dfrac{m+n}{2}\right) m^{\frac{m}{2}} n^{\frac{n}{2}} x^{\frac{m}{2}-1} (n+mx)^{-\frac{m+n}{2}}}{\Gamma\left(\dfrac{m}{2}\right)\Gamma\left(\dfrac{n}{2}\right)} & \text{se } x > 0 \\ 0 & \text{se } x \leq 0 \end{cases}$$

Caracterização

Expectância ou Valor esperado

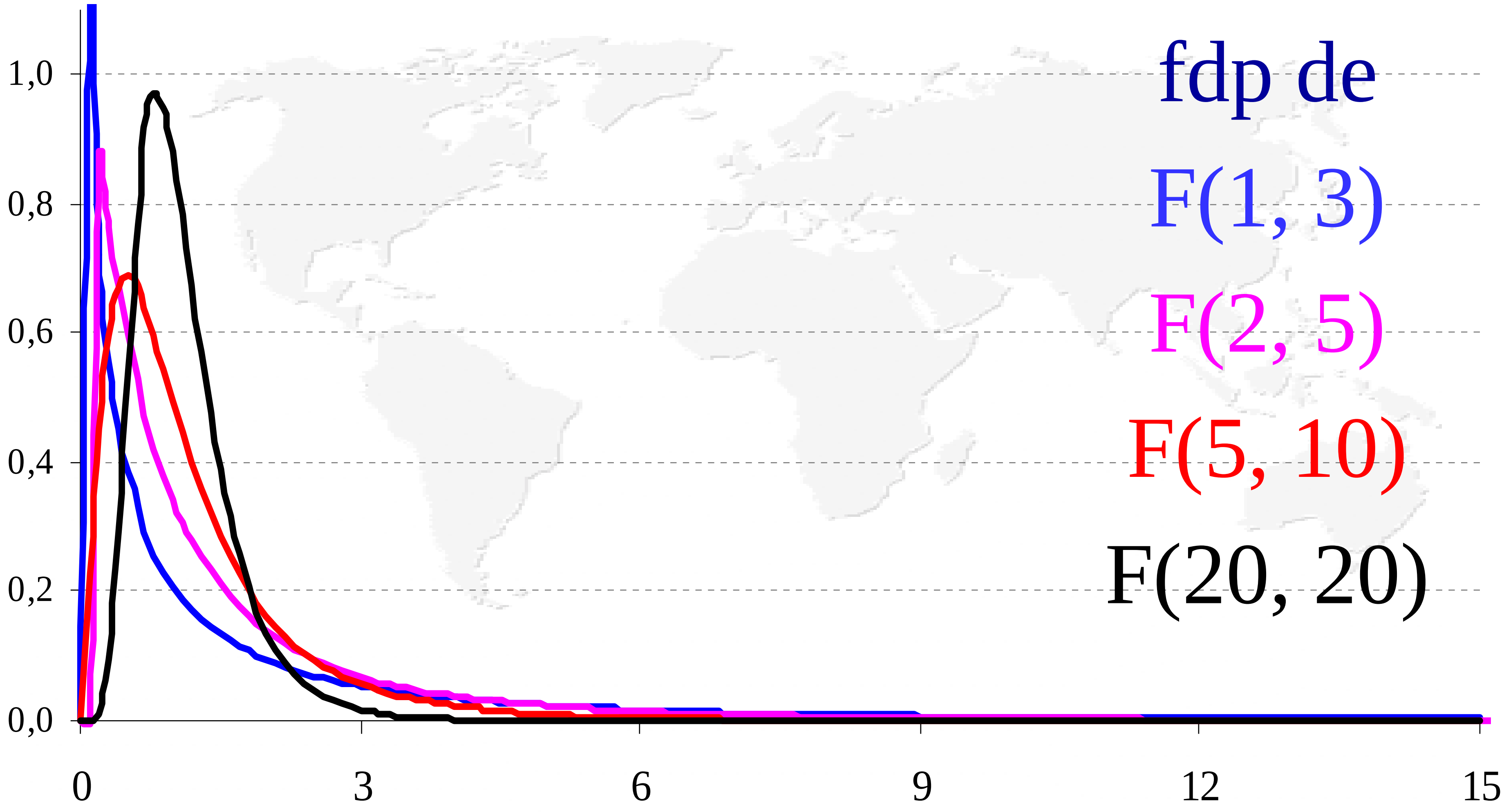
$$E(X) = \frac{m}{m-2}$$

m é o grau de liberdade do numerador e **n** do denominador

Variância

$$Var(X) = \frac{2(m+n-2)\ m^2}{m(n-2)(n-4)}$$

Gráficos



Tabelas

O que é tabelado é a percentil 95% ou 99% - área à direita de cada curva (**uma para cada par de valores – numerador, denominador**) igual a 5% e 1%, isto é, "**x**" tal que $\mathbf{P[F(m, n) \geq x] = 5\%}$ ou $\mathbf{P[F(m, n) \geq x] = 1\%}$.

	1	2	3	4	5	6	7
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77
2	18,51	[illegible]	[illegible]	[illegible]	19,30	19,33	19,35
3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	8,89
4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	6,09
5	6,61	5,79	5,41	[illegible]	5,05	4,95	4,88
6	5,99	5,14	4,76	[illegible]	4,39	4,28	4,21
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91

$P[F(5,7) \geq 3,97] = 5\%$

	1	2	3	4	5	6	7
1	4052,18	4999,34	5403,53	5624,26	5763,96	5858,95	5928,33
2	98,50	[illegible]	[illegible]	[illegible]	99,30	99,33	99,36
3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	27,67
4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	14,98
5	16,26	13,27	12,06	[illegible]	10,97	10,67	10,46
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64

$$\mathbf{P[F(5,\ 7) \geq 7{,}46] = 1\%}$$

Desigualdade de Tchebycheff

Desigualdade

de **Tchebycheff**, Tchebichev ou Chebyshev, 1821 –1894.

$$P(|X - \mu| \geq k\sigma) \leq 1/k^2$$

$$P(|X - \mu| < k\sigma) \geq 1 - 1/k^2$$

Desigualdade de Camp-Meidell

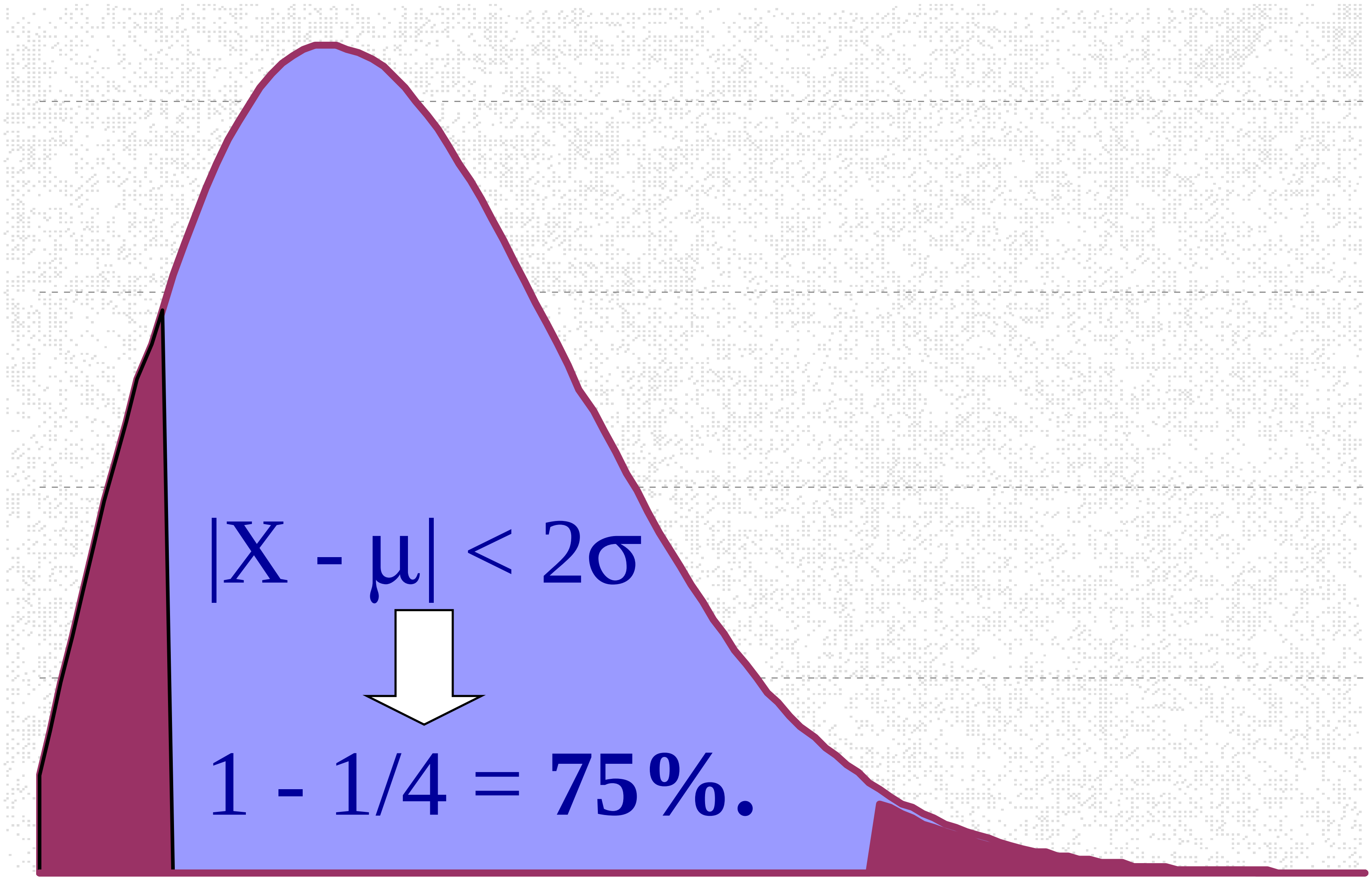
Se a distribuição for unimodal e simétrica, então:

$$P(|X - \mu| \geq k\sigma) \leq 4/9k^2$$

Estas desigualdades fornecem as probabilidades de que os valores de uma VAD/VAC estejam em um intervalo simétrico em torno da média de amplitude igual a 2k desvios padrões.

Assim se k = 2, por exemplo, a desigualdade de **Tchebycheff** estabelece que o percentual de valores da variável aleatória que está compreendida no intervalo $\mu \pm 2\sigma$ é de pelo menos $1 - 1/4 =$ **75%.**

Graficamente



Na normal este percentual vale exatamente 95,44%. Mas como a normal é simétrica e unimodal, neste caso, um resultado mais próximo é dado pela desigualdade de **Camp-Meidell**, isto é:

$$\mathbf{1 - 4/(9k^2) = 1 - (1/9) = 88{,}89\%.}$$

Exemplo

O número de aviões que chegam a um aeroporto durante um determinado de tempo tem o seguinte comportamento:

$$f(x) = \frac{100^{x} e^{-100}}{x!} \qquad x = 0, 1, 2, 3 \ldots$$

Utilize a desigualdade de **Tchebichev** para determinar uma cota inferior da probabilidade $P(85 \leq X \leq 115)$

Solução

Como k = 1,5, então a probabilidade solicitada deve ser maior ou igual a:

$$1-\frac{1}{k^2}=1-\frac{1}{1,5^2}=$$

$$1-0,4444=55,56\%$$