Distribuições de Probabilidades

Quando aplicamos a Estatística na resolução de problemas administrativos, verificamos que muitos problemas apresentam as mesmas características o que nos permite estabelecer um modelo teórico para determinação da solução de problemas.

Os componentes principais de um modelo estatístico teórico:

1. Os possíveis valores que a variável aleatória X pode assumir;
2. A função de probabilidade associada à variável aleatória X;
3. O valor esperado da variável aleatória X;
4. A variância e o desvio-padrão da variável aleatória X.

Há dois tipos de distribuições teóricas que correspondem a diferentes tipos de dados ou variáveis aleatórias: a distribuição discreta e a distribuição contínua.

Distribuições Contínuas

Variável aleatória contínua é aquela que pode assumir inúmeros valores num intervalo de números reais e é medida numa escala contínua. Por exemplo, uma variável aleatória contínua deve ser definida entre os números reais 0 e 1, ou números reais não negativos ou, para algumas distribuições, qualquer número real. A temperatura, a pressão, a precipitação ou qualquer elemento medido numa escala contínua é uma variável aleatória contínua.

Existem duas funções associadas a cada variável contínua X: a **função densidade de probabilidade**, simbolizada por **f(X)**, e a **função cumulativa de probabilidade**, ou **função de distribuição de probabilidade** representada por **F(X)**. A função **f(X)** é aquela cuja integral de X = a até X = b (b ≥ a) dá a probabilidade de que X assuma valores compreendidos no intervalo (a, b), ou seja,

$$P(a \le X \le b) = \int_a^b f(X)\, dX \qquad (1)$$

A função cumulativa de probabilidade F(b) é tal que:

$$F(b) = \Pr ob(X \le b) = \int_{-\infty}^{b} f(X)\, dX \qquad (2)$$

Qualquer função definida no campo real só pode ser considerada como uma função densidade de probabilidade se forem satisfeitas as seguintes condições:

$$f(X) \ge 0 \qquad (3)$$

para todo X e

$$F(X) = \int_{-\infty}^{\infty} X\, dX = 1 \qquad (4)$$

A probabilidade de que a variável X assuma valores no intervalo (a, b) é dada por:

$$P(a \le X \le b) = \int_a^b f(X)\, dX = F(b) - F(a) \qquad (5)$$

e a probabilidade de que a variável contínua X assuma um valor em particular, b, por exemplo, é:

$$P(a \leq X \leq b) = \int_a^b f(X)\,dX = F(b) - F(b) = 0 \qquad (6)$$

Há muitas distribuições teóricas contínuas. Algumas das mais usadas aqui são: distribuição normal, distribuição gamma, distribuição de valores extremos e distribuição exponencial. Neste material vamos tratar dos modelos probabilísticos citados, que têm importância prática na investigação científica, abordando as formas das funções densidade de probabilidade, bem como a esperança e a variância.

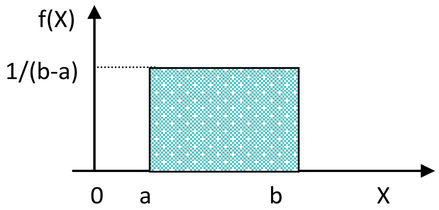
Distribuição Uniforme

Uma distribuição de variável aleatória contínua é a *distribuição uniforme* cuja função densidade de probabilidade é **constante** dentro de um intervalo de valores da variável aleatória **X.**

A variável aleatória X tem distribuição uniforme de probabilidades no intervalo (a, b) se a função densidade f(x) for:

$f(x) = \frac{1}{b-a}$, com as seguintes condições: b ≥ a e a ≤ x ≤ b.

A representação gráfica da distribuição uniforme é um retângulo com base definida pelos valores a e b que estabelecem os limites de valores possíveis da variável aleatória X, Figura XXXXX.



Da definição da *distribuição uniforme* deduzimos:

- A área do retângulo é igual a 1, pois a base é (b – a) e a altura 1/(b – a).
- A probabilidade da variável aleatória X ser igual ou maior que a e, ao mesmo tempo, menor ou igual a b é igual a 1 ou 100%

A *média* e a *variância* da variável aleatória X com *distribuição uniforme* de probabilidades no intervalo (a,b) são:

- Média: $\mu_X = \frac{a+b}{2}$
- Variância: $\sigma_X^2 = \frac{(b-a)^2}{12}$

EXEMPLO 1

A variável aleatória X tem distribuição uniforme no intervalo (50, 200). Calcular a média e o desvio padrão.

Solução

A ***média*** da variável aleatória contínua X é 150 obtida com a fórmula:

$$\mu_X = \frac{a+b}{2} = \frac{50+200}{2} = 125$$

Da mesma forma, a ***variância*** é 1875,00, obtida com a fórmula:

$$\sigma_X^2 = \frac{(b-a)^2}{12} = \frac{(200-50)^2}{12} = 1875{,}00$$

O ***desvio padrão*** é obtido como:

$$\sigma = \sqrt{1875} = 43{,}30$$

EXEMPLO 2

Continuando o Exemplo 1, qual a probabilidade de um valor da variável X se encontrar entre 110 e 150?

Solução

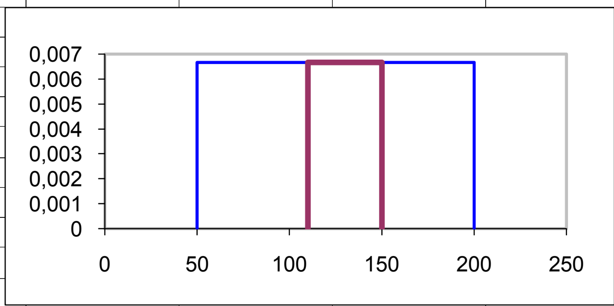
A probabilidade de um valor da variável X se encontrar entre 110 e 150 é P(110 ≤ X ≤ 150) = 0,0,2667 ou 26,67%

2. DISTRIBUIÇÃO UNIFORME NO EXCEL

O Excel não tem nenhuma função estatística para a distribuição uniforme. Entretanto é possível automatizar os cálculos criando um *modelo estatístico* para a distribuição uniforme. No segmento de planilha abaixo mostramos o modelo **Distribuição Uniforme** resolvendo os Exemplos 1 e 2. Com o modelo é possível realizar cálculos, conforme apresentado a seguir:

- As células pintadas na cor verde são células que aceitam somente dados. As células pintadas em azul são as células resultados. As restantes células pintadas de cor alaranjado são células contendo títulos.
- Nas células C4 e C5 são informados os limites *a* e *b* da variável aleatória uniforme X
- As células C6, C7 e C8 calculam, respectivamente, a média, a variância e desvio padrão.
- Informando os valores c e d pertencentes ao intervalo (*a*, *b*) nas células C10 e C11, o *modelo* calculará na célula C12 a probabilidade P(*c* ≤ X ≤ *d*). As células C10 e C11 estão preparadas para aceitar apenas valores dentro do intervalo (*a*, *b*).
- Ao mesmo tempo, o modelo constrói a função *f*(x) no intervalo (*a*, *b*) e destaca a área de cálculo da probabilidade P(*c* ≤ X ≤ *d*).

	A	B	C
1	**DISTRIBUIÇÃO UNIFORME**		
2			
3		**Variável Aleatória Uniforme *X***	
4		**Mínimo**	**50**
5		**Máximo**	**200**
6		**Média**	**125,00**
7		**Variância**	**1875,00**
8		**Desvio Padrão**	**43,30**
9		**Cálculo de Probabilidades**	
10		***c***	**110,00**
11		***d***	**150,00**
12		***P*(*c*<*X*<*d*)**	**26,67%**
13			



=SE(E(C11<=C5;C11>=C4;C10<=C5;C10>=C4);(C11-C10)/(C5-C4);"Erro!")

EXERCÍCIOS RESOLVIDOS

EXERCÍCIOS

1. Determine a probabilidade de obtermos

Distribuição Normal

1. DISTRIBUIÇÃO NORMAL – CURVA NORMAL

Entre as distribuições teóricas de variável aleatória contínua, uma das mais empregadas é a **distribuição normal.** Sua importância em análise matemática resulta do fato de que muitas técnicas estatísticas, como análise de variância, de regressão e alguns testes de hipótese, assumem e exigem a normalidade dos dados. Além disso, a ampla aplicação dessa distribuição vem em parte devido ao teorema do limite central. Este teorema declara que na medida em que o tamanho da amostra aumenta, a distribuição amostral das médias amostrais tende para uma distribuição normal (Triola, 1998).

O aspecto gráfico de uma **distribuição normal** é o da Figura 01:

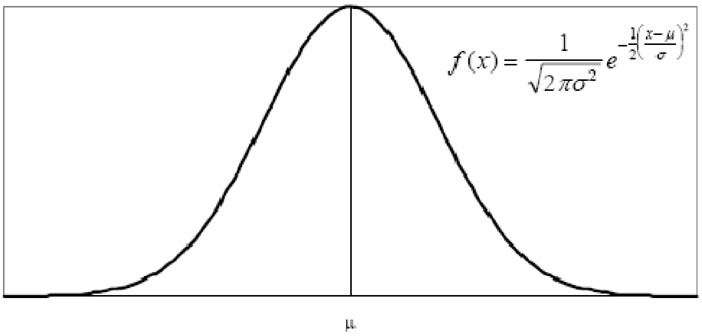


FIGURA 01

Para uma perfeita compreensão da distribuição normal, observe a Figura 01 e procure visualizar as seguintes propriedades:

1ª) A variável aleatória X pode assumir todo e qualquer valor real.

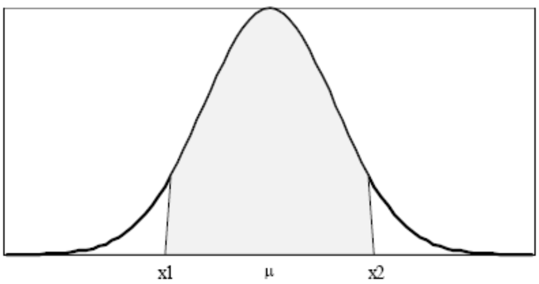
2ª) A representação gráfica da distribuição normal é uma curva em forma de sino, simétrica em torno da média (μ), que recebe o nome de **curva normal** ou de **Gauss**.

3ª) A área total limitada pela curva e pelo eixo das abcissas é igual a 1, já que essa área corresponde à probabilidade da variável aleatória X assumir qualquer valor real.

4ª) A curva normal é assintótica em relação ao eixo das abcissas, isto é, aproxima-se indefinidamente do eixo das abcissas sem, contudo, alcançá-lo.

5ª) Como a curva é simétrica em torno de μ, a probabilidade de ocorrer valor maior do que a média é igual à probabilidade de ocorrer valor menor do que a média, isto é, ambas as probabilidades são iguais a 0,5. Escrevemos: $P(X>\mu) = P(X<\mu) = 0,5$.

Quando temos em mãos uma variável aleatória com distribuição normal, nosso principal interesse é obter a probabilidade dessa variável aleatória assumir um valor em um determinado intervalo.



$$\Pr(x_1 \le x \le x_2) = \int_{x_1}^{x_2} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Calcular esta integral toda vez, não seria fácil.
A fim de ultrapassar este inconveniente, o Sr. Gauss (um dos estatísticos que inicialmente estudou esta função de distribuição) desenvolveu uma metodologia conducente à estandardização, ou redução a um caso único, de qualquer que seja a função de distribuição normal, caracterizada por μ e σ. Esta estandardização transforma qualquer função de distribuição normal $N(\mu,\sigma)$ numa única função de distribuição normal, caracterizada por ter média $\mu = 0$ e desvio padrão $\sigma = 1$, isto é, $N(0,1)$, que é designada por **função de distribuição normal reduzida.**

Vejamos como proceder, por meio de um exemplo concreto.

Seja **X** a variável aleatória que representa os diâmetros dos parafusos produzidos por certa máquina. Vamos supor que essa variável tenha distribuição normal com média μ = 2 cm desvio padrão σ = 0,04 cm.

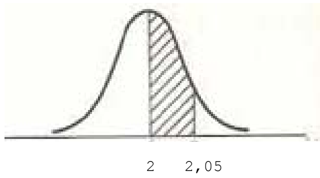
Pode haver interesse em conhecer a probabilidade de um parafuso ter um diâmetro com valor entre 2 e 2,05 cm.

É fácil notar que essa probabilidade, indicada por:

P(2 < X < 2,05),

corresponde à área hachurada na Figura 02:

Figura 02



O cálculo direto dessa probabilidade exige um conhecimento de Matemática mais avançado do que aquele que dispomos aqui. Entretanto, podemos contornar o problema facilmente. Basta aceitar, sem demonstração, que, se **X** é uma variável aleatória com distribuição normal de média μ e desvio padrão σ, então a variável:

$$\boxed{z = \frac{x - \mu}{\sigma}}$$

tem **distribuição normal reduzida**[1], isto é, tem distribuição normal de média 0 e desvio padrão 1.

As probabilidades associadas à distribuição normal padronizada são encontradas em tabelas, não havendo necessidade de serem calculadas.

A Figura abaixo é uma tabela de distribuição normal reduzida[2], que nos dá a probabilidade de **Z** tomar qualquer valor entre a média 0 e um dado valor z, isto é:

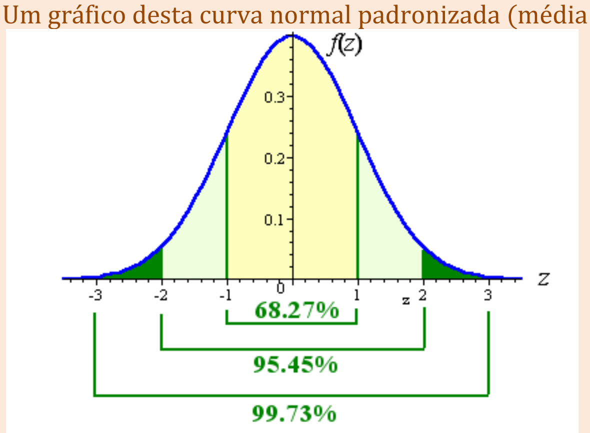
Convertidos os intervalos da **variável x**, entre os quais se pretende calcular a probabilidade, para valores padronizados em **z**, o cálculo desta probabilidade será:

$$\Pr(x_1 \le x \le x_2) = \int_{x_1}^{x_2} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \Pr(z_1 \le z \le z_2) = \int_{z_1}^{z_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

ILUSTRAÇÂO

Porcentagens da Área Sob a Curva Normal Padrão

Um gráfico desta curva normal padronizada (média 0 e Variância 1) é:



[1] O valor de Z pode ser obtido no Excel pela função **Padronizar**. Assim, =PADRONIZAR(x; média;desv_padrão)

[2] Esta Tabela foi produzida no Excel usando a função estatística =DIST.NORMP($A2+B$1)-0,5000. Foi feita a subtração de 0,5000 para os valores ficarem restritos à primeira metade dos valores acima de zero da gaussiana. Note que a função está definida assim para a célula B2. Para se obter os valores das outras células, basta arrastarmos a alça do canto inferior direito de B2 até K42.

P(0 < Z < z)

FIGURA 03 - ÁREA SUBTENDIDA PELA CURVA NORMAL REDUZIDA DE 0 A Z

Z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,10	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,20	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,30	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,40	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,50	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,60	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,70	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,80	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,90	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,00	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,10	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,20	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,30	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,40	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,50	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,60	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,70	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,80	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,90	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,00	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,10	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,20	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,30	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,40	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,50	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,60	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,70	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,80	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,90	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,00	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,10	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,20	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,30	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,40	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,50	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,60	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,70	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,80	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,90	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000
4,00	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000

Temos, então, que se **X** é uma variável aleatória com distribuição normal de média μ e desvio padrão σ, podemos escrever:

$$P(\mu < X < x) = P(0 < Z < z),$$

com $z = \frac{x-\mu}{\sigma}$.

Voltemos, então, ao nosso problema.

Queremos calcular P(2 < X < 2,05). Para obter essa probabilidade, precisamos, em primeiro lugar, calcular o valor de z que corresponde a x = 2,05 (x = 2 $\Rightarrow$ z = 0, pois μ =2). Temos, então:

$$z = \frac{x-\mu}{\sigma} = \frac{2{,}05-2}{0{,}04} = \frac{0{,}05}{0{,}04} = 1{,}25,$$

donde:

P(2 < X < 2,05) = P(0 < X 1,25)

Procuremos, agora, na Figura 03 acima o valor de z = 1,25.

Na primeira coluna encontramos o valor **1,2**. Em seguida, encontramos, na primeira linha, o valor **5**, que corresponde ao último algarismo do número **1,25**. Na intersecção da linha e coluna correspondentes encontramos o valor **0,3944**, o que nos permite escrever:

P(0 < Z < 1,25) = **0,3944**

Assim, a probabilidade de um parafuso fabricado por essa máquina apresentar um diâmetro entre a média μ = 2 e o valor x = 2,05 é **0,3944**.

Escrevemos, então:

P(2 < X < 2,05) = P(0 < Z < 1,25) = **0,3944 ou 39,44%**.

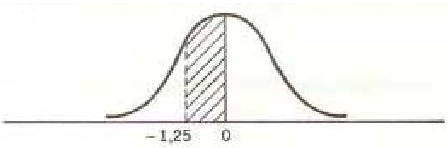
EXERCÍCIOS RESOLVIDOS

1. Determine as probabilidades:

a. P(-1,25 < Z < 0)

Solução:

A probabilidade procurada corresponde à parte hachurada da figura:



Sabemos que:

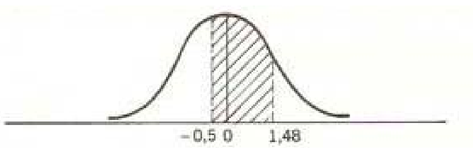
P(0 < Z < 1,25) = 0,3944

Pela simetria da curva, temos:

P(-1,25 < Z < 0 = P(0 < Z < 1,25) = 0,3944

b. P(-0,5 < Z < 1,48)

A Probabilidade procurada corresponde à parte hachurada da figura:



Temos

P(-0,5 < Z < 1,48) = P(-0,5 < Z < 0) + P(0 < Z < 1,48)

Como:

P(-0,5 < Z < 0 = P(0 < Z < 0,5) = **0,1915**

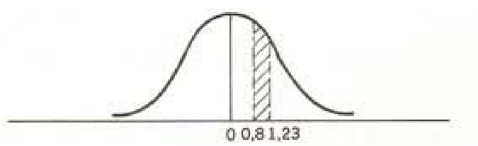
e

P(0 < Z < 1,48) = **0,4306**

obtemos:

P(-0,5 < Z < 1,48) = 0,1915 + 0,4306 = 0,6221

c. P(0,8 < Z < 1,23)



Temos:

$$P(0,8 < Z < 1,23) = P(0 < Z < 1,23) - P(0 < Z < 0,8)$$

Como:

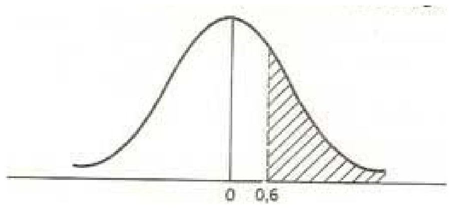
$$P(0 < Z < 1,23) = \mathbf{0,3907} \text{ e } P(0 < Z < 0,8) = \mathbf{0,2881},$$

Obtemos:

$$P(0,8 < Z < 1,23) = 0,3907 - 0,2881 = \mathbf{0,1026}.$$

d. $P(Z > 0,6)$

A probabilidade procurada corresponde à parte hachurada da figura:



Temos:

$$P(Z > 0,6) = P(Z > 0) - P(0 < Z < 0,6)$$

Como:

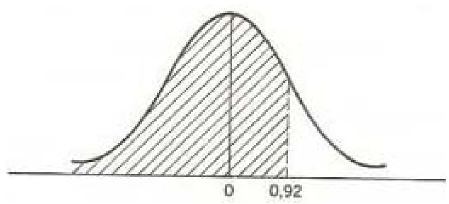
$$P(Z > 0) = \mathbf{0,5} \text{ e } P(0 < Z < 0,6) = \mathbf{0,2258}$$

obtemos:

$$P(Z > 0,6) = 0,5 - 0,2258 = \mathbf{0,2742}$$

e. $P(Z < 0,92)$

A probabilidade procurada corresponde à parte hachurada da figura:



Temos:

$$P(Z < 0,92) = P(Z < 0) + P(0 < Z < 0,92)$$

como:

$$P(Z < 0) = 0,5 \text{ e } P(0 < Z < 0,92) = \mathbf{0,3212}$$

obtemos:

$$P(Z < 0,92) = 0,5 + 0,3212 = \mathbf{0,8212}$$

2. A unidade de ensacamento de uma fábrica de cimentos é pressuposto encher os sacos com um peso médio μ=50 kg. É óbvio que nem todos os sacos ficam exatamente com a quantidade de 50 kg, havendo alguns que ficam com mais, outros que ficam com menos cimento, devido a diversos fatores aleatórios que ocasionam variabilidade no processo.

Estudada esta variabilidade ou dispersão, quantificou-se a variância do processo, tendo-se concluído que é de $\sigma^2 = 0.25$ kg^2 ou o desvio padrão $\sigma = \sqrt{0,25} = 0,5$ kg.

Admitindo que o processo de ensacamento segue a lei de distribuição normal com média $\mu = 50$ e variância $\sigma = 0.5$ (isto é, $x \sim N(\mu = 50, \sigma = 0.5)$), calcule a probabilidade de que um saco, selecionado aleatoriamente, contenha:

a) entre 50 kg e 51 kg.

b) entre 49,5 kg e 50 kg.

c) entre 49 kg e 51 kg.

d) acima de 51,5 kg.

e) abaixo de 48,75 kg.

f) entre 50,5 kg e 51,5 kg.

g) entre 48,5 kg e 49,5 kg.

h) abaixo de 48,5 kg ou acima de 51,5 kg.

i) Em 1 000 sacos saídos desta unidade de ensacamento, quantos serão esperados com o peso entre 49,5 kg e 51,5 kg?

j) Calcule os limites, inferior e superior, do intervalo central onde existem 90% dos sacos saídos desta linha de ensacamento.

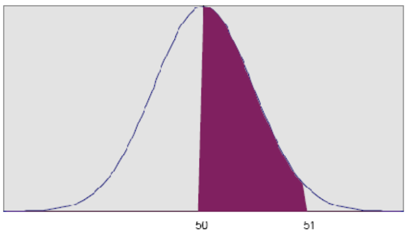
Solução:

Estabeleça-se que:

x: peso dos sacos (variável aleatória)

$x \sim N(\mu=50,\ \sigma^2=0,25)$ $\qquad \mu = 50 \qquad \sigma = 0,5$

a) Pretende-se calcular **Pr(50 ≤ x ≤ 51)**. Esta probabilidade é graficamente traduzida pela seguinte área:



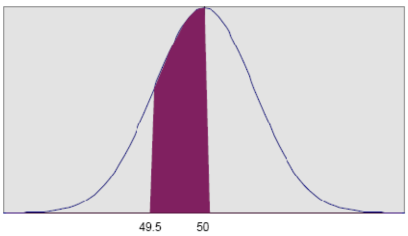
Convertam-se os limites do intervalo para a variável z normal reduzida:

- para x=50 vem: $z = \frac{x-\mu}{\sigma} = \frac{50-50}{0,5} = 0$
- para x=51 vem: $z = \frac{x-\mu}{\sigma} = \frac{51-50}{0,5} = 2$

Então:

Pr(50 ≤ x ≤ 51) = Pr(0 ≤ z ≤ 2) = 0,4772 ou 47,72%, fazendo esta leitura na tabela para z = 2,00.

b) Pretende-se calcular **Pr(49,5 ≤ x ≤ 50)**. Esta probabilidade é graficamente traduzida pela seguinte área:



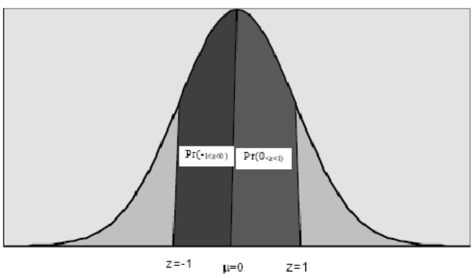
Convertam-se os limites do intervalo para a variável z normal reduzida:

- para x=49,5 vem: $z = \frac{x-\mu}{\sigma} = \frac{49,5-50}{0,5} = -1$
- para x=50 vem: $z = \frac{x-\mu}{\sigma} = \frac{50-50}{0,5} = 0$

Então:

Pr(49,5 ≤ x ≤ 50) = Pr(-1 ≤ z ≤ 0)

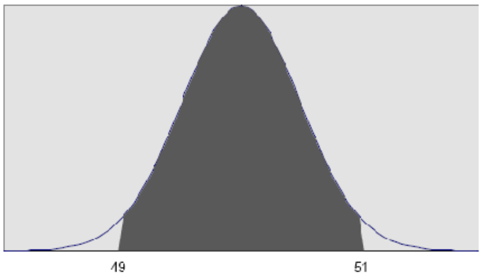
Neste ponto, depara-se à dificuldade de que a tabela anexa apenas dá os valores das probabilidades para intervalos acima de z = 0, isto é Pr(0 ≤ z ≤ z_α), sendo $z_\alpha \geq 0$. Apelando para a propriedade da simetria da distribuição normal, conclui-se que as duas áreas abaixo indicadas são idênticas, isto é, Pr(-1 ≤ z ≤ 0) = Pr(0 ≤ z ≤1).



Então:

Pr(49,5 ≤ x ≤50) = Pr(-1 ≤ z ≤0) = Pr(0 ≤ z ≤ 1) = 0,3413 ou 34,13%

c) Pretende-se calcular **Pr(49 ≤ x ≤ 51)**. Esta probabilidade é graficamente traduzida pela seguinte área:



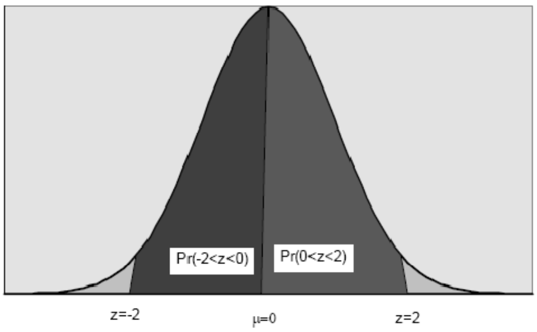
Convertam-se os limites do intervalo para a variável z normal reduzida:

- para x=49 vem: $z = \frac{x-\mu}{\sigma} = \frac{49-50}{0,5} = -2$
- para x=51 vem: $z = \frac{x-\mu}{\sigma} = \frac{51-50}{0,5} = 2$

Então:

Pr(49 ≤ x ≤ 51) = Pr(-2 ≤ z ≤ 2)

Analisando a área que traduz a probabilidade, nota-se que essa área é composta por duas partes, nomeadamente a área compreendida entre z = -2 e z = 0, no ramo inferior da curva, e pela área delimitada z = 0 e z = 2, no ramo superior. Em termos de probabilidade, tem-se:

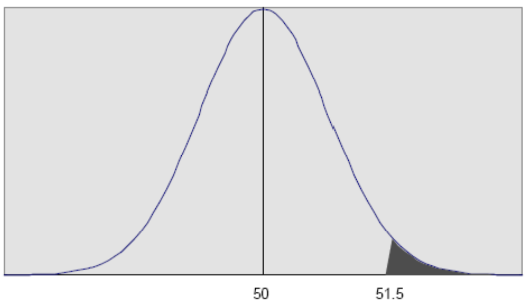


Pr(-2 ≤ z ≤ 2) = Pr(-2 ≤ z ≤ 0) + Pr(0 ≤ z ≤ 2)

Como a tabela só permite a leitura direta de Pr(0 ≤ z ≤ 2), há que transformar, pela propriedade da simetria, a área abaixo de z = 0 numa área equivalente acima de z = 0.
Fica então:

Pr(49 ≤ x ≤ 51) = Pr(-2 ≤ z ≤ 2) = Pr(-2 ≤ z ≤ 0) + Pr(0 ≤ z ≤ 2) =
(propriedade da simetria)
= Pr(0 ≤ z ≤ 2) + Pr(0 ≤ z ≤ 2) = 2 x Pr(0 ≤ z ≤ 2) = 2 x 0,4772 = 0,9544 ou 95,44%

d) Pretende-se calcular **Pr(x ≥ 51,5)**. Esta probabilidade é graficamente traduzida pela seguinte área:



Convertam-se os limites do intervalo para a variável z normal reduzida:

- para x= 51,5 vem: $z = \frac{x-\mu}{\sigma} = \frac{51,5-50}{0,5} = 3$

Então:

Pr(x ≥ 51) = Prz ≥ 2)

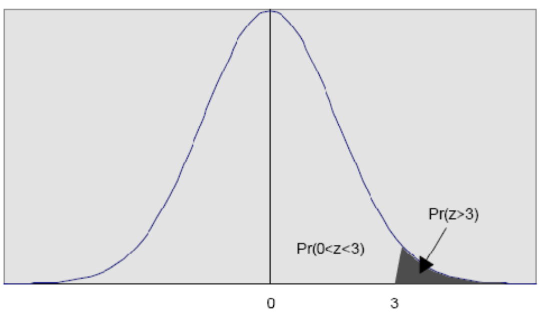
Analisando a área que traduz a probabilidade, nota-se que essa área é a cauda superior da área total à direita de z = 0, delimitada inferiormente por z = 3.

Contudo, a tabela em uso dá leituras para áreas delimitadas inferiormente por z = 0 e superiormente por z = z_α (neste caso z = 3).

Numa situação deste gênero, há que apelar para uma propriedade fundamental da distribuição normal, que estabelece que Pr(x ≥ μ) = Pr(z ≥ 0) = 0,5.

Pela análise das áreas envolvidas, depreende-se que:

Pr(z ≥ 3) = Pr(z ≥0) - Pr(0 ≤ z ≤ 3)

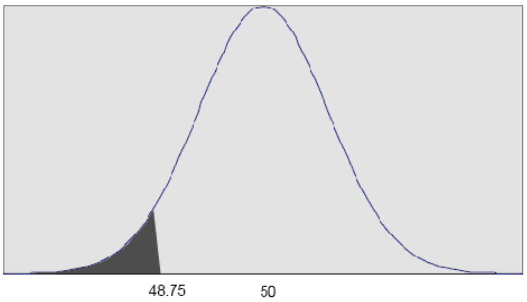


Então:

Pr(x ≥ 51,5) = Pr(z ≥ 0) – Pr(0 ≤ z ≤ 3) =

0,5 – 0,4987 = 0,0013 ou 0,13%

e) Pretende-se calcular **Pr(x ≤ 48,75).** Esta probabilidade é graficamente traduzida pela seguinte área:



Convertam-se os limites do intervalo para a variável z normal reduzida:

- para x= 48,755 vem: $z = \frac{x-\mu}{\sigma} = \frac{48{,}75-50}{0{,}5} = -2{,}5$

Aplicando a propriedade da simetria:
Pr(x ≤ 48,75) = Pr(z ≤ -2,5)= Pr(z ≥ 2,5)

Pelo que foi exposto na alínea anterior, conclui-se que:

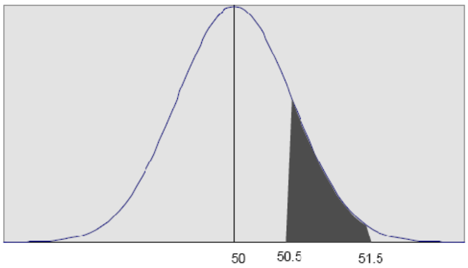
Pr(z ≤ 2,5) = Pr(z ≥ 0) - Pr(0 ≤ z ≤ 2,5)

Então:

Pr(x ≤ 48,75) = Pr(z ≤ -2,5) = Pr(z ≥ 2,5) = Pr(z ≥0) - Pr(0 ≤ z ≤ 2,5)

= 0,5 – 0,4938 = 0,0062 ou 0,62%

f) Pretende-se calcular **Pr(50,5 ≤ x ≤ 51,5).** Esta probabilidade é graficamente traduzida pela seguinte área:



Convertam-se os limites do intervalo para a variável z normal reduzida:

- para x=50,5 vem: $z = \frac{x-\mu}{\sigma} = \frac{50{,}5-50}{0{,}5} = 1$
- para x=51,5 vem: $z = \frac{x-\mu}{\sigma} = \frac{51{,}5-50}{0{,}5} = 3$

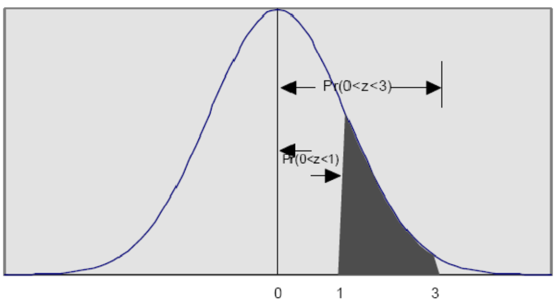
Então:
Pr(50,5 ≤ x ≤ 51,5) = Pr(1 ≤ z ≤ 3)

Note-se que a área que traduz esta probabilidade é uma área no ramo superior da curva da distribuição normal, sem que contudo os seus limites coincidam com z = 0.

Analisando as área envolvidas, conclui-se que:

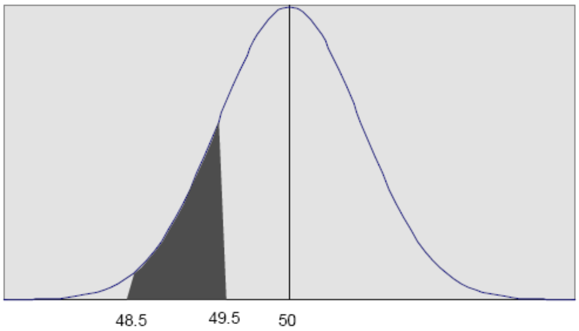
Pr(1 ≤ z ≤ 3) = Pr(0 ≤ z ≤ 3) - Pr(0 ≤ z ≤ 1)

Isto é, expressou-se a área a calcular em função da diferença de duas áreas cuja leitura é direta na tabela.



Então:
Pr(50,5 ≤ x ≤ 51,5) = Pr(1 ≤ z ≤3) = Pr(0≤z≤3) - Pr(0 ≤ z ≤1)
= 0,4987 – 0,3413 = 0,1574 ou 15,74%

g) Pretende-se calcular **Pr(48,5 ≤ x ≤ 49,5).** Esta probabilidade é graficamente traduzida pela seguinte área:



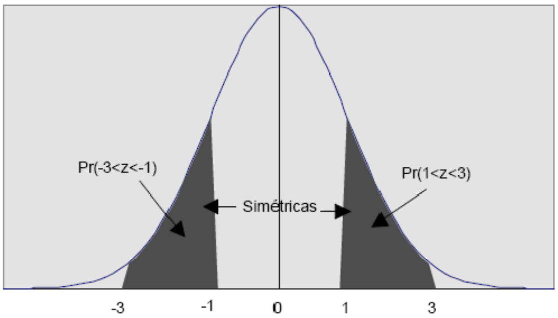
Convertam-se os limites do intervalo para a variável z normal reduzida:

- para x=48,5 vem: $z = \frac{x-\mu}{\sigma} = \frac{48{,}5-50}{0{,}5} = -3$
- para x=49,5 vem: $z = \frac{x-\mu}{\sigma} = \frac{49{,}5-50}{0{,}5} = -1$

Então:

Pr(48,5 ≤ x ≤ 49,5) = Pr(-3 ≤ z ≤ -1)

Aplicando a propriedade da simetria da distribuição normal, vem uma situação análoga à resolvida na alínea anterior:

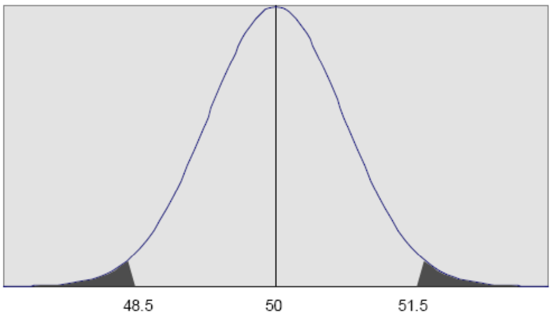


Então:

Pr(48,5 ≤ x ≤ 49,5) = Pr(-3 ≤ z ≤-1) = Pr(1≤z≤3) - Pr(0 ≤ z ≤3)- Pr(0 ≤ z ≤ 1)

= 0,4987 – 0,3413 = 0,1574 ou 15,74%

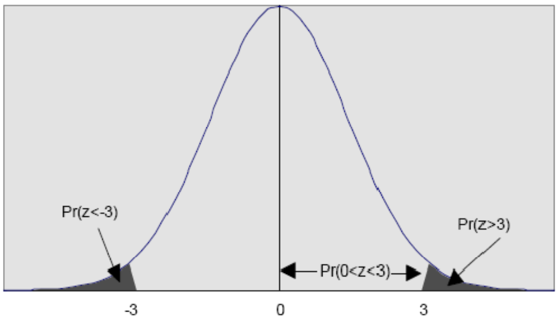
h) Pretende-se calcular **Pr(x ≤ 48,5 ∪ x ≥ 51,5)**. Esta probabilidade é graficamente traduzida pela seguinte área:



Após fazer a transformação para a curva normal N(0,1), vem:

Pr(x ≤ 48,5 ∪ x ≥ 51,5) = Pr(z ≤ -3 ∪ z ≥ 3)

Analisando as áreas envolvidas, conclui-se que:



Pr(z ≤ -3 ∪ z ≥3) = Pr(z ≤-3) + Pr(z ≥3)

Aplicando a propriedade da simetria à área que traduz a Pr(z ≤ -3), conclui-se: que:

Pr(z ≤ -3 ∪ z ≥3) = Pr(z ≤-3) + Pr(z ≥3) =

Pr(z ≥ 3) + Pr(z ≥ 3) =

= 2 x Pr(z ≥3) =

(aplicando a resolução da alínea d)

= 2 x [Pr(z ≥ 0) - Pr(0 ≤ z ≤3)] =

= 2 x [0,5 – 0,4987] =

= 2 x 0,0013 = 0,0026

i) No fundo, pretende-se calcular a proporção de sacos com peso x ∈[49,5, 50,5].

Aplicando o mesmo método de resolução da alínea c), conclui-se que:

Pr(49,5 ≤ x ≤ 50,5) = Pr(-1 ≤ z ≤1) = Pr(-1 ≤ z ≤0) + Pr(0 ≤ z ≤1) =

(propriedade da simetria)

= Pr(0 ≤ z ≤1) + Pr(0 ≤ z ≤1) =

= 2 x Pr(0 ≤ z ≤ 1) =

= 2 x 0,3413 = 0,6826

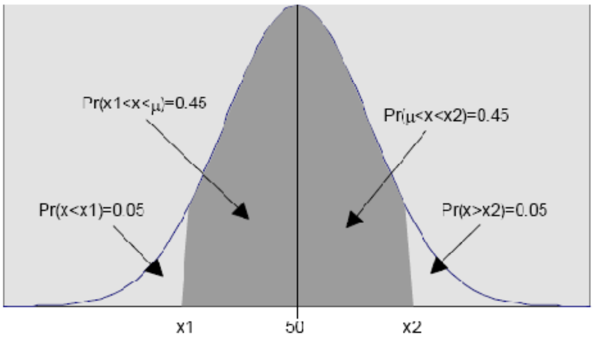
Então:

N° esperado de sacos com peso $x \in [49{,}5 , 50{,}5]$ =

= N° total de sacos x $Pr(49{,}5 \le x \le 50{,}5)$ = 1 000 x 0,6826 ≈ 683 sacos.

j) Pretendem-se calcular os limites **inferior** ($\mathbf{x_1}$) e **superior** ($\mathbf{x_2}$) do **intervalo central** onde existem 90% dos sacos saídos desta linha de ensacamento.

Graficamente, tem-se a seguinte situação, onde se sabem as seguintes probabilidades, pela análise do intervalo pretendido em conjugação com as propriedades da distribuição normal:



$Pr(x > \mu) = Pr(x < \mu) = 0{,}5$

$Pr(x_1 < x < x_2) = 0{,}90$ *(dado do enunciado)*

$Pr(x_1 < x < \mu) = 0{,}45 = Pr(\mu < x < x_2)$ *(intervalo central)*

$Pr(x < x_1) = 0{,}05 = 0{,}5 - Pr(x1 < x < \mu)$ *(propriedade da distribuição normal)*

$Pr(x > x_2) = 0{,}05 = 0{,}5 - Pr(\mu < x < x_2)$ *(propriedade da distribuição normal)*

Tendo em atenção que x_1 e x_2 são simétricos em torno de $\mu=50$ (porque definem um intervalo central), então z_1 e z_2 (redução de x_1 e x_2 respectivamente, através da expressão $z = \frac{x-\mu}{\sigma}$ são simétricos em relação a $z = 0$.

Por $Pr(\mu < x < x_2) = 0{,}45$, sabe-se que $Pr(0 < z < z_2) = 0{,}45$. Por leitura na tabela da distribuição normal, fica-se a saber que:

para $Pr(0 < z < z_\alpha) = 0{,}4495 \therefore z_\alpha = 1{,}64$

para $Pr(0 < z < z_\alpha) = 0{,}4505 \therefore z_\alpha = 1{,}65$

Como $Pr(0 < z < z_2) = 0{,}45$ está exatamente ao centro entre $Pr(0 < z < z_\alpha) = 0{,}4495$ e $Pr(0 < z < z_\alpha) = 0{,}4505$, fazendo interpolação direta nos dois valores z_α calculados anteriormente, conclui-se que $z_2 = 1{,}645$.

Então, utilizando a expressão de redução $z = \frac{x-\mu}{\sigma}$, obtém-se:

$x_2 = \mu + z_2.\sigma \therefore x_2 = 50 + 1{,}645 \text{ x } 0{,}5 \therefore x_2 = 50{,}8225$

e

$x_2 = \mu + z_2.\sigma \therefore x_2 = 50 - 1.645 \text{ x } 0.5 \therefore x_1 = 49{,}1775$ *(porque x_1 e x_2 são simétricos em torno de μ)*

Concluindo, o intervalo pretendido é: $x \in [49{,}1775 , 50{,}8225]$.

3. Os salários mensais dos executivos de uma determinada indústria são distribuídos normalmente, em torno da média de R$ 10.000, com desvio padrão de R$ 800. Calcule a probabilidade de um executivo ter um salário semanal situado entre R$ 9.800 e R$ 10.400

Solução:

Devemos, inicialmente, determinar os valores da variável de distribuição normal reduzida. Assim:

$$z_1 = \frac{9.800-10.000}{800} = -0{,}25 \text{ e } z_2 = \frac{10.400-10.000}{800} = 0{,}5$$

Logo, a probabilidade procurada é dada por:

$P(9.800 < Z < 10.400) = P(-0{,}25 < Z < 0{,}5) = P(-0{,}25 < Z < 08) + P(0 < Z < 0{,}5) =$ **0,0987 + 0,1915 = 0,2902**

É, pois, de se esperar que, em média, 29,02% dos executivos tenham salários entre R$ 9.800 e R$ 10.400.

EXERCÍCIOS

1. Sendo Z uma variável com distribuição normal reduzida, calcule:

a. $P(0 < Z < 1{,}44)$ e. $P(Z > -2{,}03)$

b. $P(-0{,}85 < Z < 05)$ f. $P(Z > 1{,}08)$

c. $P(-1{,}48 < Z < 2{,}05)$ g. $P(Z < -0{,}66)$

d. $P(0{,}72 < Z < 1{,}89)$ h. $P(Z < 0{,}60)$

2. Um teste padronizado de escolaridade tem distribuição normal com média 100 e desvio padrão 10. Determine a probabilidade de um indivíduo submetido ao teste ter nota:

a. maior que 120;

b. maior que 80;

c. entre 85 e 115;

d. maior que 100.

3. Os pesos de 600 estudantes são normalmente distribuídos com média 65,3 kg e desvio padrão 5,5 kg. Determine o número de estudantes que pesam:

a. entre 60 e 70 kg;

b. mais que 63,2 kg;

c. menos que 68 kg.

4. A duração de um certo componente eletrônico tem média de 850 dias e desvio padrão de 40 dias. Sabendo que a duração é normalmente distribuída, calcule a probabilidade desse componente durar:

a. entre 700 e 1.000 dias;

b. mais de 800 dias;

c. menos de 750 dias.

RESPOSTAS:

1. a. 0,4251 b. 0,3023 c. 0,9104 d. 0,2064 e. 0,9788 f. 0,1401 g. 0,2546 h. 0,7258

2. a. 0,0228 b. 0,9772 c. 0,8664 d. 0,5

3. a. 0,6338 b. 0,6480 c. 0,6879

4. a. 0,9998 b. 0,8944 c. 0,0062

2. PARÂMETROS DA DISTRIBUIÇÃO NORMAL

A distribuição normal é uma distribuição de dois parâmetros μ (média) e σ (desvio-padrão). A densidade de probabilidade desta distribuição tem a seguinte forma:

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(X-\mu)^2}{2\sigma^2}} \quad \text{para } -\infty < x < +\infty \tag{10}$$

onde μ e σ são a média e o desvio-padrão da população, respectivamente. O μ é estimado por $\overline{X}$ e σ por s, que são obtidos através das relações:

$$\overline{X} = \frac{\sum_{i=1}^{N} X_i}{N} \tag{11}$$

$$s^2 = \frac{\sum_{i=1}^{N} \left(X_i - \overline{X}\right)^2}{N-1} \tag{12}$$

Uma notação bastante empregada para designar que uma variável tem distribuição normal com média $\overline{X}$ e variância s^2 (s é a representação de σ e $\overline{X}$ de μ de uma amostra) é $N(\overline{X}, s^2)$. Se uma amostra de dados tem realmente distribuição normal a seguinte relação é válida: A = (K-3) = 0. A *curtose* da distribuição normal é igual a 3 e a *assimetria* é nula.

O histograma de freqüências da distribuição normal tem a forma de sino ou parecida. Com a média constante e a variância variável, o gráfico da curva normal assume diferentes formas de sino: de alongada a achatada.A probabilidade de que X assuma valores menores ou iguais a um dado x quando X é N($\overline{X}$,s^2) é estimada por:

$$F(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{X} e^{-\frac{(X-\mu)^2}{2\sigma^2}} dX \qquad (13)$$

Mas essa equação não pode ser resolvida analiticamente sem o uso de métodos de integração aproximada. Por essa razão usa-se a transformação $Z = \frac{(X - \overline{X})}{s}$ e com isso a variável Z tem N(0,1).

A variável Z é chamada variável reduzida e a curva

$$F(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Z} e^{-\frac{Z^2}{2}} dZ \qquad (14)$$

é a curva normal reduzida.

F(Z) na forma da equação (14) é tabulada. Como a curva normal reduzida é simétrica, essa propriedade é geralmente utilizada na tabulação de apenas valores positivos de Z. Mas algumas tabelas, como a tabela 4, também mostram valores negativos de Z. As tabelas de F(Z) tanto podem indicar a Prob($Z \le z$), bem como as Prob($0 \le Z \le z$). Por isso, a escolha da tabela e sua utilização deve ser feita com muito cuidado. A tabela utilizada aqui fornece Prob($Z \le z$). Mas nas tabelas que fornecem apenas os valores positivos da variável reduzida faz-se uso da propriedade de simetria da curva normal reduzida de modo que: $P(-X \le Z \le 0) = P(0 \le Z \le X)$.

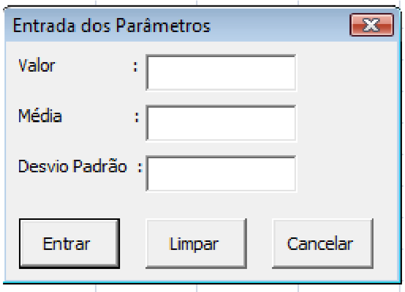
3. DISTRIBUIÇÃO NORMAL NO EXCEL

Poderíamos construir uma planilha para realizar os cálculos de $P(0 < Z < z)$ diretamente na planilha Excel. Para tanto,basta construir uma planilha como a mostrada abaixo:

	A	B	C	D	E	F
1	valor	9,00				
2	média	7,00				
3	desvio padrão	1,20				
4	valor reduzido z	1,67	<--=PADRONIZAR(B1;B2;B3)			
5	arredondado	1,60	<--=ARREDONDAR.PARA.BAIXO(B4;1)			
6	centésimo	0,07	<--B4-B5			
7	P(0<Z<z)	0,4515	<--=PROCV(B5;Dist_Z!$A$2:$K$42;B6*100+2)			
8						

O processo fica eficiente e com redução de erros de leitura por parte do usuário da Tabela. Ainda mais este resultado poderá ser utilizado em outras células de cálculo de novas variáveis que necessitam do conhecimento do valor de $P(0 < Z < z)$.

É possível automatizamos esta busca na Tabela ainda mais usando o VBA com formulários. Para isso construa o *formulário* seguinte:



O formulário Entrada dos Parâmetros é um simples formulário ilustrando os princípios de design de *UserForm* e a codificação *VBA* associada.

Ele usa uma seleção de controles onde temos três rótulos: Valor, Média e Desvio Padrão. Três caixas de textos: txtValor, txtMedia e txtDesvPad para as entradas dos parâmetros. E, ainda, três botões: BtnEntrar, BtnLimpar e BtnCancelar. Quando o usuário clicar o botão Entrar suas entradas são lançadas nas células correspondentes na planilha.

As configurações das propriedades dos controles são:

Controle	Tipo	Propriedade	Configuração
UserForm	UserForm	Name	*frmEntradaParametros*
		Caption	*Entrada dos Parâmetros*
Valor	Text Box	Name	*txtValor*
Média	TextBox	Name	*txtMedia*
Desvio Padrão	TextBox	Name	*txtDesvPad*
Entrar	Command Button	Name	*BtnEntrar*
		Caption	*Entrar*
		Default	*True*
Limpar	CommandButton	Name	*BtnLimpar*
		Caption	*Limpar*
		Default	*True*
Cancelar	CommandButton	Name	*BtnCancelar*
		Caption	*Cancelar*
		Default	*True*

CONSTRUÇÃO DO FORMULÁRIO

Se você quiser construir este formulário, simplesmente copie o layout mostrado na ilustração acima. Siga os passos abaixo:

1. Abra a pasta (*workbook*) que você quer que o formulário pertença (*UserForms* como macros tem de serem atribuídos a uma pasta) e ligue o VBE do Excel.

2. No VBE clique no botão Inserir *UserForm* (ou vá para *Inserir > UserForm*)

3. Se a caixa de ferramentas não aparecer por si só (primeiro clique no *form* para garantir-se que ele não está oculto) clique no botão Caixa de Ferramentas.

4. Para colocar um controle no formulário clique no botão apropriado na caixa de ferramentas e daí clique no formulário. Controles podem ser movidos arrastando-os pelos seus lados, ou redimensionando arrastando os botões ao redor do perímetro.

5. Para editar as propriedades de um controle, certifique-se que o controle escolhido esteja selecionado e daí faça as mudanças apropriadas na janela *Properties*. Se você não puder ver a janela *properties*, vá para **Exibir > Janela de Propriedades.**

6. Para remover um controle de um formulário, selecione-o e clique a tecla *Delete* no seu teclado.

Um *UserForm* realmente não fará qualquer coisa até o código que dirige o formulário e seus vários controles seja criado. O próximo passo é escrever o código que dirige o próprio formulário.

Inicializando o Formulário:

A maioria dos formulários precisa de uma espécie de configuração quando são abertos. Neles podem ser definidos valores default, certifique-se de que os campos estejam vazios. Este processo é chamado de inicialização do formulário e ele é tratado por uma macro chamada **UserForm_Initialize.** Aqui está como construir o código para inicializar o Formulário de Entrada dos Parâmetros:

1. Para ver a janela de código do formulário vá para **Exibir > Código** ou clique **F7.**

2. Quando a janela de código se abrir primeiramente ela conterá um procedimento **UserForm_Click( )** vazio. Usamos as listas *drop-down* no topo da janela de código para escolher **UserForm** e **Initialize.** Isto criará o procedimento que você precisa. Você pode agora deletar o procedimento **UserForm_Click( ).**

```
Private Sub UserForm_Initialize()
    txtValor.Value = ""
    txtMedia.Value = ""
    txtDesvPad.Value = ""
    txtValor.SetFocus
End Sub
```

O propósito do procedimento **UserForm_Initialize( )** é preparar o formulário para uso, configurando os valores default para os vários controles.

As linhas:

```
txtValor.Value = ""
txtMedia.Value = ""
txtDesvPad.Value = ""
```

definem os conteúdos das duas caixas de texto para vazio.

A linha:

```
txtValor.SetFocus
```

coloca o cursor do usuário na caixa de texto *txtValor* de modo que ela não precisa ser clicada antes de começar a digitar.

Existem três botões de comendo no formulário e cada um deve ser potencializado pelo seu próprio procedimento. Comecemos com o mais simples deles, o botão Cancelar.

Anteriormente, usamos a Janela *Properties* para definir a propriedade *Cancel* do botão Cancelar para *True*. Quando você configurar a propriedade Cancelar de um botão de comando para *True*, esta tem o efeito de “clicar” aquele botão quando o usuário pressionar a tecla *Esc* no seu teclado. Mas ela sozinha não fará qualquer coisa acontecer para o formulário. Você precisa criar o código para o evento clique do botão que fechará, neste caso, o formulário. Aqui está como:

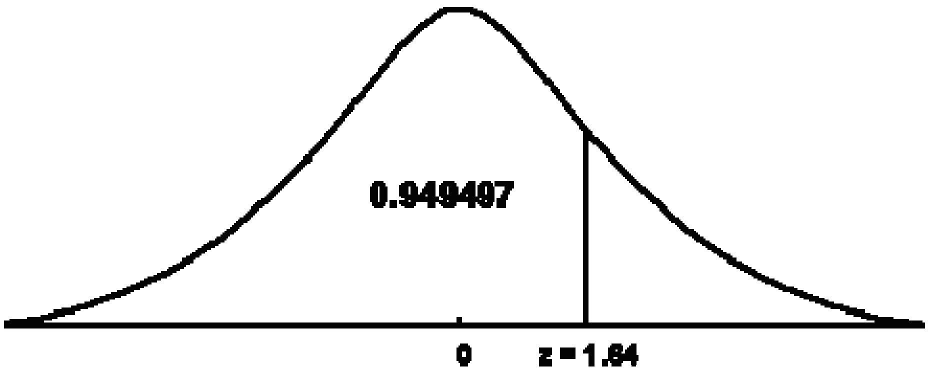
1. Com o *UserForm* aberto

MACRO (FUNÇÃO) PARA DISTRIBUIÇÃO NORMAL

Distribuição Normal Padrão Acumulada

Esta função calcula a area sob o lado esquerdo de um valor especificado (o valor z) de uma curva de função densidade de distribuição normal padrão (standard normal distribution density function curve). Num português simples, ela retorna a probabilidade de X que é menor que um valor específico.

Se você não souber com o que uma curva normal se parece ou já se esqueceu dela, aqui está um exemplo:



Neste exemplo, a probabilidade de X ser menor que 1,64 (z) é 94.9497

```
Function u_SNorm(z)

    c1 = 2.506628
    c2 = 0.3193815
    c3 = -0.3565638
    c4 = 1.7814779
    c5 = -1.821256
    c6 = 1.3302744
    If z > 0 Or z = 0 Then
            w = 1
    Else: w = -1
    End If
    y = 1 / (1 + 0.2316419 * w * z)
    u_SNorm = 0.5 + w * (0.5 - (Exp(-z * z / 2) / c1) * _
          (y * (c2 + y * (c3 + y * (c4 + y * (c5 + y * c6))))))

End Function


u_SNorm(1.64) = 0.949497
```

Esta função também é implementada no exemplo Black-Scholes Option Pricing Model - European Call and Put.

(Esta função é similar à função **NORMSDIST()** fornecida pelo Excel.)

4. APLICAÇÃO – O Mercado de Ações

Algumas vezes, os mercados de ações seguem uma tendência para cima (ou tendência para baixo) dentro de 2 desvios padrões da média. Isto é chamado mover-se dentro do **canal de regressão linear.**

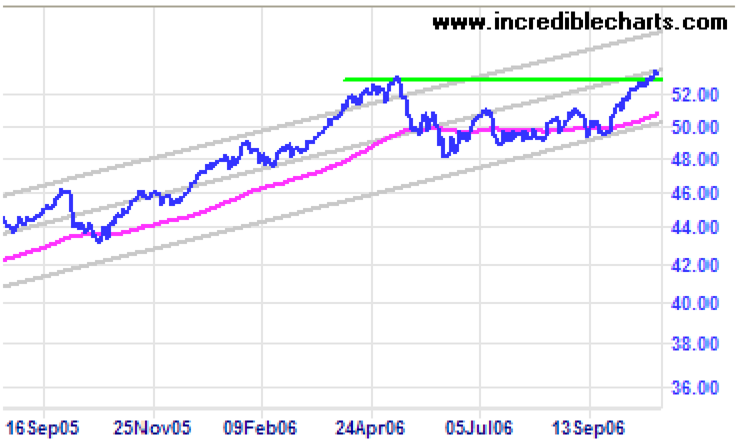
Aqui está um gráfico do Australian index (o All Ordinaries) de 2003 até Set 2006.

Fonte da imagem: incrediblecharts.com.

A linha cinza superior está 2 desvios padrões acima da media e a linha cinza inferior está 2 desvios padrões abaixo da média.

Note que em Abril de 2006 o índice esteve acima d a margem superior do canal e uma correção seguida (o mercado despencou).

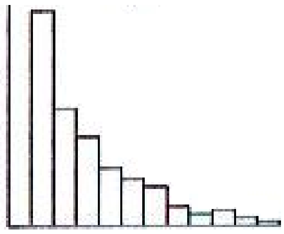
Mas de forma interessante, a última parte do gráfico mostra que o índice somente esteve em queda até o ponto no fundo do canal e daí então recuperou até a média, como você pode ver na exibição ampliada abaixo. Tais análises ajudam os traders ganharem dinheiro (ou não perderem dinheiro) quando estão investindo.



Fonte da imagem: incrediblecharts.com.

Distribuição Exponencial

A distribuição exponencial é geralmente aplicada à dados com forte assimetria[3] como aqueles cujo histograma tem a forma da figura abaixo, ou seja, de J invertido. Quando os serviços prestados por uma empresa para clientes externos ou internos são de duração variável é esta distribuição a indicada para analisar esses experimentos, por exemplo, a duração do atendimento do caixa de um banco ou de postos de saúde, o tempo de operação sem interrupção de um equipamento, etc. Sua **densidade de probabilidade** tem a forma:



$$f(x) = \lambda e^{-\lambda X} \quad \text{com } \lambda > 0, x \geq 0 \tag{1}$$

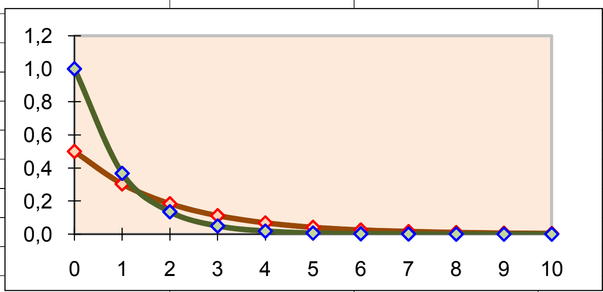
e sua **função de distribuição de probabilidade** é do tipo:

$$F(x) = \int_0^{\infty} \lambda e^{-\lambda X} = 1 - e^{-\lambda X} \tag{2}$$

As características da função exponencial definida são:

- A distribuição não é simétrica como mostra a Figura abaixo para dois valores do parâmetro λ, obtida no segmento de planilha:

	A	B	C	D	E	F	G	H
1	DISTRIBUIÇÃO EXPONENCIAL							
2								
3		λ	0,5	1				
4		Média	2,00	1,00				
5		Desvio Padrão	2,00	1,00				
6		X	f(X)	f(X)				
7		0	0,5000	1,0000				
8		1	0,3033	0,3679				
9		2	0,1839	0,1353				
10		3	0,1116	0,0498				
11		4	0,0677	0,0183				
12		5	0,0410	0,0067				
13		6	0,0249	0,0025				
14		7	0,0151	0,0009				
15		8	0,0092	0,0003				
16		9	0,0056	0,0001				
17		10	0,0034	0,0000				
18								



=C$3*EXP(-$B17*C$3)

- A variável aleatória X assume somente valores positivos.
- Comparando com a *distribuição normal*, enquanto esta é completamente definida por dois parâmetros, *média* e *desvio padrão*, a **distribuição exponencial** é definida por apenas um único parâmetro λ, estimado por:

$$\lambda = \frac{1}{\mu} \tag{3}$$

Com isso, a **função cumulativa de probabilidade** assume a forma geralmente encontrada na literatura, ou seja:

$$F(X) = 1 - e^{-\frac{a}{\mu}} \quad \text{para } 0 \leq X \leq a \quad \text{e}$$
$$F(X) = e^{-\frac{a}{\mu}} \quad \text{para } X \geq a \tag{4}$$

A *esperança* e a *variância* da **distribuição exponencial** são obtidas através das expressões: $\mu = 1/\lambda$ e $\sigma^2 = 1/\lambda^2$, respectivamente. A distribuição exponencial é um caso especial da distribuição gama com o parâmetro $\lambda = 1$.

[3] Skewness

EXEMPLO 1 (Projeto PAE – Bolsista: Michelle S. Reboita)

Considere os dados diários de chuva de Pelotas – RS, no mês de *janeiro*, cuja distribuição de freqüências consta na tabela 15. Neste exemplo os dados brutos não são apresentados.

Os cálculos necessários para a estimativa da *média* e da *variância* dos dados também estão indicados na tabela 15, com isso, tem-se:

$$\sum f = 450 + 184 + 80 + 43 + 23 + 9 + 7 + 5 + 2 + 2 + 0 + 1 = 806$$

$$\sum fX = 450 \times 5{,}5 + 184 \times 15 + 80 \times 25 + 43 \times 35 + 23 \times 45 + 9 \times 55 + 7 \times 65 + 5 \times 75 + 2 \times 85 + 2 \times 95 + 0 \times 105 + 1 \times 115 = 11575$$

$$\sum fX^2 = 450 \times 5{,}5^2 + 184 \times 15^2 + 80 \times 25^2 + 43 \times 35^2 + 23 \times 45^2 + 9 \times 55^2 + 7 \times 65^2 + 5 \times 75^2 + 2 \times 85^2 + 2 \times 95^2 + 0 \times 105^2 + 1 \times 115^2 = 334912{,}5$$

Tabela 15. Distribuição de freqüências dos totais diários de chuva de janeiro de Pelotas, RS, no período de 1893 a 1994. Foram considerados apenas os valores > 1,0 mm.

Classes	PM (X)	f	f . X	f . X²	F(X)	fe
1 – 10	5,5	450	2475	13612,5	0,5016	404
10 – 20	15	184	2760	41400,0	0,7516	201
20 – 30	25	80	2000	50000,0	0,8762	100
30 - 40	35	43	1505	52675,0	0,9383	50
40 – 50	45	23	1035	46575,0	0,9692	25
50 - 60	55	9	495	27225,0	0,9847	12
60 -70	65	7	455	29575,0	0,9924	6
70 – 80	75	5	375	28125,0	0,9962	3
80 – 90	85	2	170	14450,0	0,9981	2
90 -100	95	2	190	18050,0	0,9990	1
100 – 110	105	0	0	0,0	0,9995	0
110 -120	115	1	115	13225,0	0,9998	0
Totais	-	806	11575	334912,5	-	806

$$\bar{X} = \frac{\sum fX}{\sum f} = \frac{11575}{806} = 14{,}361$$

$$s^2 = \frac{\left|\sum fX^2 - \left(\sum fX\right)^2 / \sum f\right|}{\sum f - 1} = \frac{334912{,}5 - 11575^2 / 806}{805} = 209{,}54$$

s = 14,48

$$\lambda = \frac{1}{\mu} = \frac{1}{14{,}361} = 0{,}0696$$

Os valores de F(X) e as freqüências **esperadas** são assim calculados:

$F(X_1)$ = 1-exp(-0,0696 x 10) =0,5016 ⇒ fe = 404

$F(X_2)$ = 1-exp(-0,0696 x 20) =0,7516 ⇒ fe = 201

$F(X_3)$ = 1-exp(-0,0696 x 30) =0,8762 ⇒ fe = 100

$F(X_4)$ = 1-exp(-0,0696 x 40) =0,9383 ⇒ fe = 50

$F(X_5)$ = 1-exp(-0,0696 x 50) =0,9692 ⇒ fe = 25

$F(X_6)$ = 1-exp(-0,0696 x 60) =0,9847 ⇒ fe = 12

$F(X_7)$ = 1-exp(-0,0696 x 70) =0,9924 ⇒ fe = 6

$F(X_8)$ = 1-exp(-0,0696 x 80) =0,9962 ⇒ fe = 3

$F(X_9)$ = 1-exp(-0,0696 x 90) =0,9981 ⇒ fe = 2

$F(X_{10})$ = 1-exp(-0,0696 x 100) =0,9990 ⇒ fe = 1

$F(X_{11})$ = 1-exp(-0,0696 x 110) =0,9995 ⇒ fe = 0

$F(X_{12})$ = 1-exp(-0,0696 x 120) =0,9998 ⇒ fe = 0

O histograma dos dados da tabela 15 está apresentado abaixo:

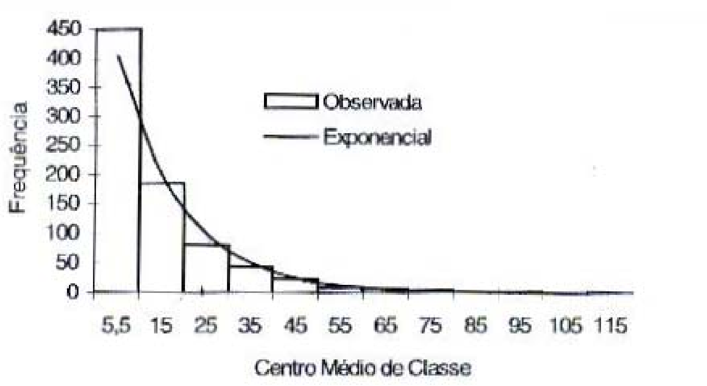


Figura 8. Distribuição exponencial ajustada aos totais diários de chuva de janeiro de Piracicaba – SP, no período de 1917 a 1989 (Assis et al., 1996, pg. 72).

EXEMPLO 2

O prazo de operação medido em horas de uma máquina de embalagem de frascos sem interrupções para manutenção tem distribuição exponencial com média de 2 horas. Qual a probabilidade desta máquina conseguir operar mais de 1 hora sem interrupção?

Solução

A probabilidade da máquina de embalagem de frascos em conseguir operar 1 hora ou mais sem interrupção é P(X ≥ 1). Da distribuição exponencial acumulada, complementar, com média de 2 horas e λ = 0,50 obtemos 60,65 com a fórmula $P(x \geq 1) = e^{-\frac{a}{\mu}} = e^{-\frac{1}{\frac{1}{0,50}}} = 0,6065$ ou 60,65%

3. DISTRIBUIÇÃO EXPONENCIAL NO EXCEL

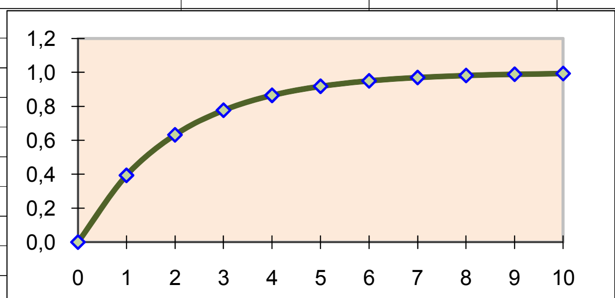
Para a distribuição exponencial, o Excel dispõe da função estatística DISTEXPON cuja sintaxe é:

DISTEXPON(x;lambda;cumulativo)

Que dá a função densidade de x ou a probabilidade acumulada de zero até x, conforme o argumento *cumulativo*.

- Se cumulativo for FALSO, a função estatística DISTEXPON dá a função densidade $f(x) = \lambda\, e^{-\lambda X}$, considerando o parâmetro lambda. Esta função está mostrada no segmento abaixo, onde podemos escolher na caixa de combinação o tipo de cumulativo (verdadeiro ou falso):

	A	B	C	D	E	F	G
22		Função DISTEXPON					
23		λ	0,5				
24		*X*	VERDADEIRO				
25		0	0,0000				
26		1	0,3935				
27		2	0,6321				
28		3	0,7769				
29		4	0,8647				
30		5	0,9179				
31		6	0,9502				
32		7	0,9698				
33		8	0,9817				
34		9	0,9889				
35		10	0,9933				
36							



- Se cumulativo for VERDADEIRO, a função DISTEXPON dá a probabilidade acumulada de zero até x, P(0 ≤ X ≤ x) considerando o parâmetro lambada, valor obtido com a fórmula $P(X \leq a) = 1 - e^{-\frac{a}{\mu}}$. Por exemplo, a probabilidade acumulada do exemplo 2 pode ser obtida com a fórmula: =1-DISTEXPON(1;0,5;VERDADEIRO) → 0,6065. Escolhendo na caixa de combinação VERDADEIRO a planilha constrói a curva de probabilidade acumulada..

Distribuição Log-Normal

Nem todas as variáveis aleatórias têm distribuição normal. Há experiências com resultados **não simétricos**, por exemplo, o retorno das operações financeiras.

A variável aleatória X com valores positivos tem *distribuição log-normal* com função densidade de probabilidade:

$$\begin{cases} f(x) = \dfrac{1}{x\sigma_Y\sqrt{2\pi}} e^{-\frac{1}{2}x\frac{(\ln x - \mu_Y)^2}{\sigma_Y^2}} & para\ x \geq 0 \\ f(x) = 0 & para\ x < 0 \end{cases}$$

se a variável aleatória Y definida como Y = *ln*(X) tiver distribuição normal com *média* $-\infty < \mu_Y < +\infty$ e *desvio padrão* $0 \leq \sigma_Y < \infty$.

Analisando a variável aleatória X retorno de um investimento em ações:

- A relação entre o resgate e a aplicação pode ser maior que 1, sem nenhuma limitação até onde o próprio mercado permitir.
- Entretanto, a relação entre o resgate e a aplicação pode ser menor que 1 até o limite de não resgatar nada e perder a aplicação realizada, provocando uma distribuição de retornos assimétrica.

A *média* e a *variância* de X com *distribuição log-normal* são:

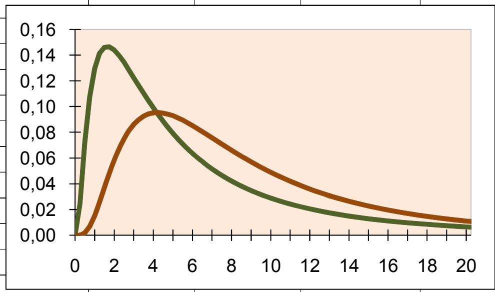
$$\mu_X = e^{\mu_X + \frac{\sigma_Y^2}{2}}$$

$$\sigma_X^2 = e^{2\mu_Y + \sigma_Y^2}\ x\ (e^{\sigma_Y^2} - 1)$$

3. DISTRIBUIÇÃO Log-Normal NO EXCEL

Vejamos o segmento de planilha com esta distribuição:

	A	B	C	D	E	F	G	H	I
1	DISTRIBUIÇÃO LOG-NORMAL								
2									
3		Parâmetros da Distribuição Normal *Y*							
4		μ_Y	1,5	2					
5		σ_Y	1	0,75					
6		Distribuição Log-normal *X*							
7		μ_X	7,39	9,79					
8		σ_X	9,69	8,51					
9		Função densidade							
10		Intervalo da curva:		0,25					
11		*x*	*f*(*x*)	*f*(*x*)					
12		0,0	0,0000	0,0000					
13		0,3	0,0248	0,0001					
14		0,5	0,0720	0,0017					
15		0,8	0,1076	0,0068					
16		1,0	0,1295	0,0152					
17		1,3	0,1412	0,0257					
18		1,5	0,1461	0,0370					
19		1,8	0,1465	0,0481					
20		2,0	0,1441	0,0583					
21		2,3	0,1398	0,0673					
22		2,5	0,1346	0,0749					
23		2,8	0,1288	0,0812					
24		3,0	0,1227	0,0861					
25		3,3	0,1166	0,0899					
26		3,5	0,1106	0,0925					
27		3,8	0,1047	0,0942					
28		4,0	0,0991	0,0951					
29		4,3	0,0937	0,0954					
30		4,5	0,0887	0,0950					
31		4,8	0,0838	0,0941					
32		5,0	0,0793	0,0929					
33		5,3	0,0750	0,0913					



Variando a média e o desvio padrão, células C4 e D5, da distribuição normal Y = ln(X) pode-se analisar o comportamento destas curvas. No intervalo de células C7:D8 a planilha fornece a média e o desvio padrão de cada distribuição log-normal, como mostrado na figura acima.

O Excel dispõe das funções estatísticas DIST.LOGNORMAL e INVLOG para cálculos com a distribuição log-normal.

A sintaxe da função DIST.LOGNORMAL é:

DIST.LOGNORMAL(x;média;desv_padrão)

A função DIST.LOGNORMAL dá a probabilidade acumulada de 0 a x, conhecidos os argumentos *média* e *desv_padrão*.

Veja um exemplo:

	J	K	L	M	N	O
1	Função DIST.LOGNORMAL					
2						
3		μ_Y	1,5			
4		σ_Y	1			
5		*x*	4			
6		*P(X<=4)*	0,4547	<--=DIST.LOGNORMAL(L5;L3;L4		
7		*P(X<=4)*	0,4547	<--=DIST.NORMP((LN(L5)-L3)/L4		
8						

A sintaxe da função INVLOG é:

INVLOG(probabilidade;média;desv_padrão)

A função INVLOG dá o valor de x para a probabilidade, conhecidos os argumentos *média* e *desv_padrão*. Em outras palavras, a função INVLOG é a função inversa da função DIST.LOGNORMAL.

Veja um exemplo:

	J	K	L	M	N	O	P
9	Função INVLOG						
10							
11		μ_Y	1,5				
12		σ_Y	1				
13		*Probabilidade*	0,4547				
14		*x*	4,00	<--=INVLOG(L13;L11;L12)			
15		x	4,00	<--=EXP(L11+L12*INV.NORMP(L13))			
16							

Como a distribuição log-normal é relacionada com a distribuição normal, a probabilidade acumulada de zero até x na distribuição log-normal com parâmetros μ e σ é igual à probabilidade acumulada de -∞ até ln(x) da distribuição normal com média μ e desvio padrão σ; isto é:

$$P(X \leq x) = DIST.LOGNORMAL(x, \mu, \sigma) = DIST.NORM\left(\frac{\ln(x) - \mu}{\sigma}\right)$$

Esta igualdade pode ser verificada nas células L7 e L8 da planilha acima. Da mesma maneira, o cálculo de x para uma determinada probabilidade acumulada considerando os parâmetros da distribuição log-normal tem a seguinte equivalência com a distribuição normal:

$$x = INVLOG(p, \mu, \sigma) = e^{[\mu + \sigma\ x\ INV.NORM(p)]}$$

Esta igualdade pode ser verificada nas células L14 e L15 da planilha acima.

Distribuição Gama

Muitas variáveis aleatórias contínuas possuem assimetria (*skewness*) positiva, ou seja, são distorcidas à direita. Freqüentemente a distorção ocorre quando há um limite físico à esquerda que é relativamente próximo a variação dos dados (*Wilks*, 1995). Exemplos comuns desta situação são as quantias de precipitação e a velocidade do vento que são fisicamente não negativas. Há uma variedade de distribuições contínuas que são limitas à esquerda por zero. Entretanto, a distribuição gama é comumente usada para representar dados de precipitação.

A **função densidade de probabilidade** da distribuição gama é:

$$f(x; \propto; \beta) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} \qquad (1)$$

onde, β é um parâmetro de escala, α é o parâmetro de forma e Γ(α) é a função gama ordinária de α. A função gama tem as seguintes propriedades:

$$\Gamma(X) = \int_0^{\infty} X^{\alpha-1} e^{-X} dx \qquad (2)$$

para todo X > 0

$\Gamma(1) = \Gamma(2) = 1$

$\Gamma(X) = \Gamma(X-1)! \quad \text{para} \ \ X = 1, 2, 3, \ldots$

$\Gamma(X+1) = X\Gamma(X) \quad \text{para} : X > 0$

$\Gamma(1/5) = \sqrt{\pi} = 1{,}77245$

O valor de Γ(X) pode ser obtido, com boa aproximação, através da seguinte relação: $A = \frac{2}{\sqrt{\alpha}}$

$$\Gamma(X) = \sqrt{\frac{2\pi}{X}} e^{X[\ln(x) - f(x)]} \qquad (3)$$

onde:

$$f(X) = 1 - \frac{1}{12X^2} + \frac{1}{360X^4} - \frac{1}{1260X^6} \tag{4}$$

A tabela 7 fornece os valores de $\Gamma(X)$, com base nestas relações.

A *média*, a *variância* e o *coeficiente de assimetria* (A) da distribuição gama podem ser obtidos por:

$$\mu = \alpha\,\beta \tag{5}$$

$$\sigma^2 = \alpha\,\beta^2 \tag{6}$$

$$A = \frac{2}{\sqrt{\alpha}} \tag{7}$$

A distribuição gama tem assimetria positiva com o parâmetro β diminuindo e o parâmetro α aumentando. Variando-se β, com α constante, muda-se a escala da distribuição, enquanto variando-se α, com β constante, muda-se a sua forma.

Quando $\alpha = 1$, DISTGAMA retornará a distribuição exponencial com:

$$\lambda = \frac{1}{\beta}$$

Para um inteiro positivo n, quando $\alpha = n/2$, $\beta = 2$, e cumulativo = VERDADEIRO, a DISTGAMA retornará (1 - DIST.QUI(x)) com n graus de liberdade.

Quando α for um positivo inteiro, DISTGAMA também será chamada de distribuição *Erlang*.

Tabela GAMA. Função gama de Y.

Y	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0,10										
0,20	42,580	46,461	50,778	55,578	60,917	66,854	73,458	80,806	88,982	98,083
0,30	3,901	3,856	3,815	3,778	3,743	3,712	3,683	3,657	3,634	3,613
0,40	2,347	2,311	2,277	2,245	2,214	2,185	2,157	2,131	2,105	2,081
0,50	1,802	1,778	1,755	1,733	1,712	1,692	1,673	1,654	1,636	1,619
0,60	1,498	1,482	1,466	1,451	1,436	1,422	1,408	1,395	1,383	1,370
0,70	1,301	1,290	1,278	1,268	1,257	1,247	1,238	1,228	1,219	1,211
0,80	1,166	1,157	1,149	1,142	1,134	1,127	1,120	1,113	1,107	1,101
0,90	1,069	1,063	1,058	1,052	1,047	1,042	1,037	1,032	1,028	1,023
1,00	1,000	0,996	0,992	0,988	0,985	0,981	0,978	0,975	0,972	0,969
1,10	0,952	0,949	0,946	0,944	0,941	0,939	0,937	0,935	0,933	0,932
1,20	0,918	0,917	0,915	0,914	0,913	0,911	0,910	0,909	0,909	0,908
1,30	0,898	0,897	0,896	0,896	0,896	0,895	0,895	0,895	0,895	0,895
1,40	0,887	0,888	0,888	0,888	0,889	0,889	0,890	0,891	0,892	0,892
1,50	0,886	0,887	0,888	0,889	0,891	0,892	0,893	0,895	0,896	0,898
1,60	0,894	0,895	0,897	0,899	0,901	0,903	0,905	0,907	0,910	0,912
1,70	0,909	0,911	0,914	0,916	0,919	0,922	0,924	0,927	0,930	0,933
1,80	0,931	0,935	0,938	0,941	0,945	0,948	0,952	0,955	0,959	0,963
1,90	0,962	0,966	0,970	0,974	0,978	0,982	0,986	0,991	0,995	1,000
1,99	0,996	1,000	1,005	1,010	1,015	1,020	1,025	1,030	1,035	1,040
2,00	1,000	1,005	1,009	1,014	1,019	1,024	1,029	1,035	1,040	1,045

Pode-se concluir, com base na equação (7), que, quando α tende para infinito $A \Rightarrow 0$, ou seja, a *distribuição gama*, neste caso, tende a ser *simétrica*.

As estimativas dos parâmetros β e α resultam da solução das equações (5) e (6). Mas essas estimativas não são adequadas, preferindo-se as estimativas descritas em Thom (1966):

$$\alpha = \frac{1}{4A}\left(1 + \sqrt{1 + \frac{4A}{3}}\right) \tag{8}$$

$$\beta = \frac{\mu}{\alpha} \tag{9}$$

sendo

$$A = \ln \overline{X} - X_g \tag{10}$$

onde

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i \tag{11}$$

é a média aritmética e

$$X_g = \frac{1}{N}\sum_{1}^{N} \ln(X_i) \tag{12}$$

é a média geométrica das observações, ou alternativamente, segundo Greenwood e Durand (1960) dada por:

$$\alpha = \frac{0{,}5000876 + 0{,}1648852\, Z - 0{,}054427\, Z^2}{Z} \tag{13}$$

quando $0 \le Z \le 0{,}5772$ e por

$$\alpha = \frac{8{,}898919 + 9{,}05995\, Z - 0{,}9775373\, Z^2}{Z(17{,}79728 + 11{,}968477\, Z + Z^2} \tag{14}$$

quando $0{,}5772 < Z < 7{,}0$, onde

$$Z = \ln\left(\overline{X}\right) - X_g \tag{15}$$

Neste caso o parâmetro β continua sendo calculado como na equação (23).

A **função cumulativa de probabilidade** é:

$$F(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}}\int_0^X X^{\alpha-1} e^{-\frac{X}{\beta}} dx \tag{16}$$

Esta equação não tem solução imediata, exigindo tabelas ou técnicas de integração numérica como expansão em série e a *fórmula de Simpson*, por exemplo. A série normalmente utilizada é a seguinte:

$$F(\zeta) = \frac{\zeta^{\alpha}}{\alpha\Gamma(\alpha)e^{\zeta}}\left[1 + \frac{\zeta^1}{\alpha+1} + \frac{\zeta^2}{(\alpha+1)(\alpha+2)}\right] + \cdots + \frac{\zeta^3}{(\alpha+1)(\alpha+2)(\alpha+3)} \tag{17}$$

Na equação (15), fazendo-se $\zeta = \frac{X}{\beta}$; x=βt; dx=βdt, chega-se a equação (17).

A probabilidade de ocorrer um valor de $X \le t$ é F(t).

EXEMPLO 1 (Projeto PAE – Bolsista: Michelle S. Reboita)

Considerem-se os 95 valores mensais de chuva do mês de janeiro em Pelotas, RS, na tabela 8, cuja distribuição de freqüências é mostrada na tabela 9.

Solução

Considerando-se a tabela 9, tem-se:

$$\sum f = 18 + 28 + 20 + 13 + 9 + 4 + 2 + 1 = 95$$
$$\sum fX = 18 \times 31{,}1 + 28 \times 73{,}1 + 20 \times 115{,}1 + 13 \times 157{,}1 + 9 \times 199{,}1 + 4 \times 241{,}1 + 2 \times 283{,}1 + 1 \times 325{,}1 = 10.598{,}5$$

$$\mu = \frac{10.598{,}5}{95} = 111{,}56$$

$$\sum fX^2 = 18 \times 31{,}1^2 + 28 \times 73{,}1^2 + 20 \times 115{,}1^2 + 13 \times 157{,}1^2 + 9 \times 199{,}1^2 + 4 \times 241{,}1^2 + 2 \times 283{,}1^2 + 1 \times 325{,}1^2 = 1.608.101{,}75$$

$$\sigma^2 = \frac{\left[\sum fX^2 - \frac{(\sum fX)^2}{\sum f}\right]}{(\sum f - 1)} = \frac{\left[1.608.101{,}75 - \frac{10.598{,}5^2}{95}\right]}{94} = 4.528{,}72$$

$$\sum \ln(X)\, f = 18xln(31{,}1) + 28xln(73{,}1) + 20xln(115{,}1) + 13xln(157{,}1) + 9xln(199{,}1) + 4xln(241{,}1) + 2xln(283{,}1) + 1xln(325{,}1) = 429{,}3573$$

$$A = \ln(111{,}93) - \frac{429{,}3573}{95} = 0{,}19504$$

Tabela 8. Chuva mensal de *janeiro* em Pelotas, RS, no período de 1895 a 1989.

Ano	0	1	2	3	4	5	6	7	8	9
189...						112,6	32,1	129,9	183,1	63,4
190...	68,3	77,5	113,3	35,8	145,6	22,3	20,2	15,5	121,4	148,5
191...	203,6	117,8	81,3	50,1	197,7	132,6	130,1	72,8	86,6	23,1
192...	81,5	65,7	159,0	182,0	28,8	129,6	33,4	82,7	59,3	119,7
193...	97,0	239,6	31,5	59,0	151,7	45,7	64,5	64,5	232,0	92,4
194...	269,0	271,3	68,3	25,1	244,7	44,1	113,4	101,8	340,3	87,6
195...	10,4	84,9	62,8	144,4	160,1	22,1	210,9	58,4	162,0	134,5
196...	143,5	106,6	64,5	151,1	11,5	48,1	107,8	84,4	191,3	105,2
197...	83,9	148,1	178,1	213,9	127,0	129,8	140,1	119,7	72,5	14,7
198...	59,6	85,4	71,0	135,9	246,8	78,6	166,0	82,7	149,5	209,4

Tabela 9. Distribuição de freqüências dos totais mensais de chuva de *janeiro* em Pelotas – RS. Ajuste à distribuição gama.

Classes	Ponto Médio (X)	f	f . X	f . X²	ln(X) . f
10,1 – 52,1	31,1	18	559,8	17.409,78	61,8697
52,1 – 94,1	73,1	28	2.046,8	149.621,08	120,1712
94,1 – 136,1	115,1	20	2.302,0	264.960,20	94,9160
136,1 – 178,1	157,1	13	2.042,3	320.846,33	65,7395
178,1 – 220,1	199,1	9	1.791,9	356.767,29	47,6443
220,1 – 262,1	241,1	4	964,4	232.516,84	21,9408
262,1 - 304,1	283,1	2	566,2	160.291,22	11,2916
304,1 – 346,1	325,1	1	325,1	105.609,01	5,7841
Totais	**-**	**95**	**10.598,5**	**1.608.101,75**	**429,3573**

$$\alpha = \left(\frac{1}{4}x0{,}19504\right)x\left[1 + \sqrt{1 + \frac{4x0{,}19594}{3}}\right] = 2{,}7206$$

$$\beta = \frac{111{,}56}{2{,}7206} = 41{,}0066$$

$\Gamma(\alpha) = \Gamma(2{,}7206)$ é estimada pela equação (3), na qual

$$f(\alpha) = 1 - \frac{1}{12x2{,}7206^2} + \frac{1}{360x2{,}7206^4} - \frac{1}{1260x2{,}7206^6} = 0{,}98879$$

$$\Gamma(\alpha) = \sqrt{\frac{2\pi}{\alpha}}e^{\alpha[\ln(\alpha) - f(\alpha)]} = \sqrt{\frac{2\pi}{2{,}7206}}e^{2{,}7206[\ln 2{,}7206 - 0{,}98879)]} = 1{,}5704$$

As estimativas dos parâmetros com base nas equações (5) e (6) a fim de comparações ficam:

$\mu = \alpha\,\beta = 2{,}7206 \text{ x } 41{,}0066 \cong 115{,}56$

$\sigma^2 = \alpha\,\beta^2 = 2{,}7206 \text{ x } (41{,}0066)^2 = 4.574{,}80$

Com os parâmetros β e α estimado têm-se, então, a *função densidade de probabilidade*, na forma da equação (1),

$$f(x;\propto;\beta) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}$$

$$f(x) = 2{,}61 \,.\, 10^{-5} \,.\, X^{1{,}7206} \,.\, e^{-\frac{X}{41{,}0066}}$$

e a *função cumulativa de probabilidade* (equação 16) será:

$$\mathrm{F}(X) = \ 2{,}61 \,.\, 10^{-5} \int_0^x X^{1{,}7206} e^{-\frac{X}{41{,}0066}} dx$$

A solução dessa equação exige o emprego de técnicas de integração numérica ou uso de tabelas específicas. Adotou-se aqui a expansão em série na forma da equação (17), cuja reprodução de todos os cálculos é praticamente impossível de ser apresentada aqui. Mas, considerando apenas a primeira classe da distribuição de frequências, a título de exemplo, tem-se:

$$\zeta = \frac{52{,}1}{41{,}0066} = 1{,}2705$$

$$F(\zeta) = \frac{1{,}2705^{2{,}7206}}{2{,}7206 \,.1{,}5704) e^{1{,}2705}} \left(1 + \frac{1{,}2705}{3{,}7206} + \frac{1{,}2705^2}{3{,}7206 x 4{,}7206} + \frac{1{,}2705^3}{3{,}7206 x 4{,}7206 x 5{,}7206} + \frac{1{,}2705^4}{3{,}7206 x 4{,}7206 x 5{,}7206 x 6{,}7206} + \frac{1{,}2705^5}{3{,}7206 x 4{,}7206 x 5{,}7206 x 6{,}7206 x 7{,}7206}\right)$$

$= 0{,}12602\ (1 + 0{,}341484 + 0{,}091909 + 0{,}020413 + 0{,}003859) = 0{,}12602 \text{ x } 1{,}4583.$

$F(X_1) = F(52{,}1) \cong 0{,}1838$

Os valores de F(X) e as freqüências **esperadas** são assim calculados:

$F(X_1) = F(52{,}1) = 0{,}1838 \Rightarrow fe = 17$

$F(X_2) = F(94{,}1) = 0{,}4734 \Rightarrow fe = 28$

$F(X_3) = F(136{,}1) = 0{,}7052 \Rightarrow fe = 22$

$F(X_4) = F(178{,}1) = 0{,}8490 \Rightarrow fe = 14$

$F(X_5) = F(220{,}1) = 0{,}9271 \Rightarrow fe = 7$

$F(X_6) = F(262{,}1) = 0{,}9663 \Rightarrow fe = 4$

$F(X_7) = F(304{,}1) = 0{,}9849 \Rightarrow fe = 2$

$F(X_8) = F(346{,}1) = 0{,}9934 \Rightarrow fe = 1$

Tabela 10. Distribuição de freqüências dos totais mensais de chuva de *janeiro* em Pelotas – RS, ajustados à distribuição gama de probabilidade.

Classes	Ponto Médio (X)	f	F(X)	fe
10,1 – 52,1	31,1	18	0,1838	17
52,1 – 94,1	73,1	28	0,4734	28
94,1 – 136,1	115,1	20	0,7052	22
136,1 – 178,1	157,1	13	0,8489	14
178,1 – 220,1	199,1	9	0,9272	7
220,1 – 262,1	241,1	4	0,9663	4
262,1 - 304,1	283,1	2	0,9849	2
304,1 – 346,1	325,1	1	0,9934	1
Totais	-	95	-	95

O histograma de freqüências deste exemplo é mostrado na figura 6.

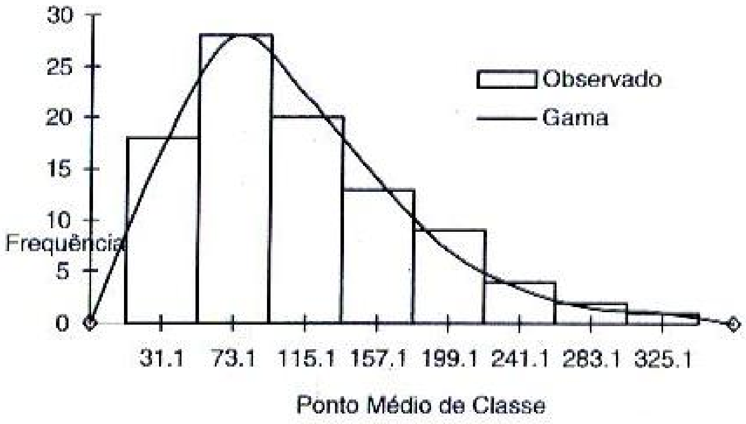


Figura 6. Totais de chuva mensal de *janeiro* em Pelotas, RS, ajustados a distribuição gama (Assis et al., 1996, pg. 59).

3. DISTRIBUIÇÃO GAMA NO EXCEL

O Excel dispõe das funções estatísticas DISTGAMA e INVGAMA para cálculos com a distribuição gama.

A sintaxe da função DISTGAMA é:

DISTGAMA(x;alfa;beta;cumulativo)

A função DISTGAMA retorna a distribuição gama, conhecidos argumentos *alfa* e *beta*, parâmetros da distribuição, números positivos. Se $\beta = 1$, a DISTGAMA retorna a distribuição *gama padrão*. O argumento *cumulativo* é um valor lógico: retornar a **função de distribuição cumulativa** = VERDADEIRO, retornar a **função de probabilidade de massa** = FALSO, ou não especificado.

A sintaxe da função INVGAMA é:

INVGAMA(probabilidade;alfa;beta)

Ela retorna o inverso da distribuição cumulativa gama. Se p = DISTGAMA(x;...), então INVGAMA(p;...) = x. Você pode usar esta função para estudar uma variável cuja distribuição pode ser enviesada.

Probabilidade é a probabilidade associada à distribuição gama.

Alfa é um parâmetro da distribuição.

Beta é um parâmetro para a distribuição. Se $\beta = 1$, INVGAMA retornará a distribuição gama padrão.

Dado um valor de probabilidade, INVGAMA procura aquele valor x de modo que DISTGAMA(x, alfa, beta, VERDADEIRO) = probabilidade. Assim, a precisão de INVGAMA depende da precisão de DISTGAMA. INVGAMA utiliza uma técnica de busca interativa. Se a busca não tiver convergido após 100 iterações, a função retornará o valor de erro #N/D.

Distribuição t de Student

De acordo com o *teorema do limite central*, a **distribuição amostral**[4] de uma estatística (como uma média da amostra) seguirá uma distribuição normal, enquanto o tamanho da amostra for suficientemente grande. Portanto, quando conhecermos o desvio padrão da população, podemos calcular um z-escore[5], e usarmos a distribuição normal para avaliar probabilidades com a média amostral.

Mas os tamanhos das amostras são algumas vezes pequenos, e frequentemente não conhecemos o desvio padrão da população. Quando um destes problemas ocorrerem, os estatísticos contam com a distribuição da *estatística t* (também conhecida como *t-escore*), cujos valores são dados por:

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

Onde $\bar{x}$ é a média amostral, μ é a média da população, s é o desvio padrão da amostra e n é o tamanho da amostra. A distribuição da *estatística t* é chamada de **distribuição t** ou de **distribuição t de Student**.

A distribuição *t de Student* tem grande importância para a inferência de parâmetros da população e para a estatística de pequenas amostras.

Graus de Liberdade

Existem realmente muitas *distribuições t* diferentes. A forma particular da *distribuição t* é determinada pelos seus graus de liberdade. Os graus de liberdade se referem ao número de observações independentes num conjunto de dados.

Quando estimar um escore médio ou uma proporção de uma amostra simples, o número de observações independentes é igual ao tamanho da amostra menos um. Daí então, a distribuição da *estatística t* das amostras de tamanho 8 serão descritas por uma distribuição t tendo 8 – 1 ou 7 graus de liberdade. Similarmente, uma *distribuição t* tendo 15 graus de liberdade seria usada com uma amostra de tamanho igual a 16.

A notação utilizada para graus de liberdade é gl[6].

Para outras aplicações, os graus de liberdade podem ser calculados diferentemente. Descreveremos estes cálculos quando eles surgirem.

Propriedades da Distribuição t

A distribuição t tem as seguintes propriedades:

- A média da distribuição é igual a 0.
- A variância é igual a $\upsilon/(\upsilon - 2)$, onde υ é o grau de liberdade (ver última seção) e $\upsilon \geq 2$.

[4] Suponha que retiremos todas as amostras possíveis de tamanho n de uma dada população. Suponha, ainda mais, que calculemos uma estatística (p.ex., uma média, desvio padrão) para cada amostra. A distribuição de probabilidade desta estatística é chamada de *distribuição amostral*.

[5] Um escore-z (também conhecido como escore padrão) indica quantos desvios padrões um elemento está da média. Um escore-z pode ser calculado pela seguinte fórmula.
$z = (X - \mu) / \sigma$
onde z é o z-escore, X é o valor do elemento, μ é a média da população, e σ é o desvio padrão.
Aqui está como interpretar os z-escores.

- Um z-escore menor que 0 representa um elemento menor que a média.
- Um z-escore maior que 0 representa um elemento maior que a média.
- Um z-escore igual a 0 representa um elemento igual à média.
- Um z-escore igual a 1 representa um elemento que está 1 desvio padrão maior que a média; um z-escore igual a 2, 2 desvios padrões maior que a média; etc.
- Um z-escore igual a -1 representa um elemento que está 1 desvio padrão menor que a média; a z-escore igual a -2, 2 desvio padrão menor que a média; etc.
- Se o número de elementos no conjunto for grande, cerca de 68% dos elementos tem um z-escore entre -1 e 1; cerca de 95% tem um z-escore entre -2 e 2; e cerca de 99% tem um z-escore entre -3 e 3.

[6] Alguns autores utilizam a notação inglesa df (degree of free)

- A variância é sempre maior que 1, embora ela esteja próxima de 1 quando existirem muitos graus de liberdade. Com infinitos graus de liberdade, a distribuição t é a mesma que a distribuição normal padrão.

Quando Usar a Distribuição t

A distribuição t pode ser usada com qualquer estatística tendo uma distribuição com a forma de sino (isto é, aproximadamente normal). O teorema do limite central estabelece que a *distribuição amostral* de uma estatística será normal ou aproximadamente normal, se qualquer uma das condições seguinte se aplicar:

- A *distribuição da população* é normal.
- A *distribuição amostral* é simétrica, unimodal, sem *outliers*, e o tamanho da amostra está entre 15 ou menos
- A *distribuição amostral* é moderadamente assimétrica, unimodal, sem *outliers*, e o tamanho da amostra está entre 16 e 40.
- O tamanho da amostra é maior que 40, sem *outliers*.

A *distribuição t* não deverá ser usada com amostras pequenas das populações que não forem aproximadamente normais.

Probabilidade e a Distribuição t de Student

Quando uma amostra de tamanho n for extraída de uma população tendo uma distribuição normal (ou aproximadamente normal), a média amostral pode ser transformada numa *t-escore*, usando a equação apresentada no início da lição. Repetimos aquela equação abaixo:

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

Onde $\bar{x}$ é a média amostral, μ é a média da população, s é o desvio padrão da amostra, n é o tamanho da amostra e os graus de liberdade são iguais a n – 1.

A *t-escore* produzida por esta transformação pode ser associada com uma única probabilidade cumulativa. Esta probabilidade cumulativa representa a probabilidade de se encontrar uma média amostral menor que ou igual a $\bar{x}$, dada uma amostra aleatória de tamanho n.

Função de Probabilidade

A função densidade de probabilidade é dada por:

$$f(t,\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}} \cdot \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

Onde Γ é a função Gama e t $\in \Re$.

A média é dada por E(t) = 0 e a variância $\mathrm{Var}(t) = \frac{\nu}{\nu-2}$.

A Distribuição t de Student no Excel

O Excel dispõe das funções estatísticas DISTT e INVT para a distribuição t cujas sintaxes são as seguintes:

DISTT(t;graus_liberdade;caudas)

A função estatística DISTT dá a probabilidade do valor t ser excedido considerando os argumentos graus-liberdade e caudas da distribuição t

- Se o argumento caudas for igual a 1, a função DISTT dará a probabilidade correspondente a uma cauda da distribuição.
- Se o argumento caudas for igual a 2, a função DISTT dará a probabilidade correspondente às duas caudas da distribuição.

INVT(probabilidade;graus_liberdade)

A função estatística INVT dá o *t-crítico* da distribuição t referente aos argumentos probabilidade e graus_liberdade, considerando que a probabilidade se refere às duas caudas da distribuição. A função INVT é a função inversa da DISTT quando o argumento caudas é igual a 2. Para o cálculo da função INVT o Excel aplica um procedimento iterativo até alcançar um erro de $\pm 3\mathrm{x}10^{-7}$. Se em 100 iterações não for possível obter o resultado, a função INVT apresenta #N/A.

A planilha abaixo mostra como podemos usar as funções da distribuição t para um exemplo.

	A	B	C	D	E	F
1	Distribuição t Student					
2						
3		Função DISTT			Uma cauda	
4					Duas caudas	
5		*t*	**1,896**		2	
6		*n*	**40**			
7			Duas caudas			
8		*g.l.*	**39**			
9		*P(t >1,896)*	**0,065**	<--=DISTT(C5;C8;E5)		
10		="P(t >"&C5&")"				
11						
12		Função INVT				
13						
14		*P*	**0,065**			
15		*n*	**40**			
16			Duas caudas			
17		*g.l.*	**39**	<--=C15-1		
18		*t*	**1,896**	<--=INVT(C14*SE(E5=1;2;1);C17)		
19						

Nesta planilha foram construídos dois modelos:

- O primeiro modelo calcula a probabilidade considerando a escolha realizada na caixa de combinação: Duas caudas ou Uma cauda, resultados previstos na própria função DISTT.
- O segundo modelo calcula o t-crítico considerando a escolha realizada na caixa de combinação: Duas caudas ou Uma cauda. O resultado para uma cauda não está previsto na função INVT; entretanto, como o valor do argumento probabilidade deverá ser o dobro do valor do problema, na célula C18 registramos a fórmula: =INVT(C14*SE(E5=1;2;1);C17) sendo E5 o endereço da célula vinculada com a caixa de combinação Duas caudas ou Uma cauda.

Notação e t escore

Os estatísticos usam t_α para representar a t-escore que tem uma distribuição de probabilidades cumulativa de (1 - α). Por exemplo, suponha que estamos interessados no t-escore tendo uma probabilidade cumulativa de 0,95. Neste exemplo, α será igual a (1 – 0,95) ou 0,05. Referiremos ao t-escore como $t_{0,05}$.

É claro, o valor de $t_{0,05}$ depende do número de graus de liberdade. Por exemplo, com 2 graus de liberdade, aquele $t_{0,05}$ é igual a 2,92; mas com 20 graus de liberdade, aquele $t_{0,05}$ é igual a 1,725.

Nota: Devido a distribuição t ser simétrica ao redor de uma média zero, o seguinte é verdadeiro:

$t_\alpha = -t_{1-\alpha}$ e $t_{1-\alpha} = -t_\alpha$

Assim, se $t_{0,05} = 2,92$, então $t_{0,95} = -2,92$.

Testando o seu entendimento

EXEMPLO 1

A Tomaz Edison fabrica lâmpadas incandescentes. O CEO exige que uma lâmpada da TE sobreviva em média 300 dias. Um pesquisador seleciona aleatoriamente 15 lâmpadas para teste. As lâmpadas amostradas sobreviveram em média 290 dias, com um desvio padrão de 50 dias. Se a exigência do CEO for verdadeira, qual é a probabilidade que 15 lâmpadas selecionadas aleatoriamente teriam uma vida média de não mais que 290 dias?

Solução

A primeira coisa que precisamos fazer é calcular o t-escore, baseado na seguinte equação:

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

$$t = \frac{290 - 300}{\frac{50}{\sqrt{15}}} = \frac{-10}{12{,}909945} = -0{,}7745966$$

Onde $\bar{x}$ é a média amostral, μ é a média da população, s é o desvio padrão da amostra e n é o tamanho da amostra.

Agora, estamos prontos a usar a planilha acima para os cálculos:

	A	B	C
1	**Distribuição t Student**		
2			
3		**Função DISTT**	
4			
5		*t*	**0,7745966**
6		*n*	**14**
7			Uma cauda
8		*g.l.*	**13**
9		*P(t >0,7745966)*	**0,226**

A planilha encontrou a probabilidade cumulativa: 0,226. Portanto, se a vida verdadeira da lâmpada fosse 300 dias, há uma chance de 22,6% que a vida média da lâmpada para 15 lâmpadas selecionadas aleatoriamente será menor que ou igual a 290 dias.

EXEMPLO 2

Suponha os escores de um teste de QI estejam normalmente distribuídos, com média de 100. Suponha que 20 pessoas sejam selecionadas aleatoriamente e testadas. O desvio padrão no grupo amostral é 15. Qual é a probabilidade que a média do escore do teste no grupo amostral será no máximo 110?

Solução

Graus de liberdade – gl = 20 – 1 = 19

Média da população = 100

Média da amostra = 110

Desvio padrão da amostra = 15

$$t = \frac{110 - 100}{\frac{15}{\sqrt{20}}} = \frac{10}{3{,}354101966} = 2{,}98142397$$

Entrando comeste valor na planilha como aquela acima, temos:

	A	B	C
1	**Distribuição t Student**		
2			
3		**Função DISTT**	
4			
5		*t*	**2,98142397**
6		*n*	**20**
7			Uma cauda
8		*g.l.*	**19**
9		*P(t >2,98142397)*	**0,004**

A planilha encontrou a probabilidade: 0,0046. Portanto, a probabilidade cumulativa é 0,996, ou seja, há 99,6% de chance que a média amostral não será maior que 110.

Distribuição Qui-Quadrado

Suponha que conduzimos o seguinte experimento estatístico. Selecionamos uma amostra aleatória de tamanho n de uma população normal, tendo um desvio padrão igual a σ. Encontramos que o desvio padrão da nossa amostra é igual a s. Com estes dados, definimos uma estatística, chamada **qui-quadrado**, usando a seguinte equação:

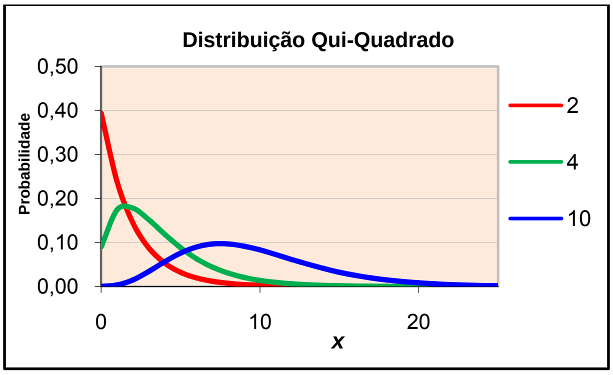
$$\chi^2 = \frac{[(n-1) * s^2]}{\sigma^2}$$

Se repetirmos este experimento um número infinito de vezes, poderemos obter uma distribuição amostral para a estatística qui-quadrado. A **distribuição qui-quadrado** é definida pela seguinte *função de densidade de probabilidade* (fdp):

$$Y = Y_0 * (\chi^2)^{\left(\frac{\nu}{2}-1\right)} * e^{-\frac{\chi^2}{2}}$$

Onde Y_0 é uma constante que depende do número de graus de liberdade, χ^2 é a estatística qui-quadrado, ν = n-1 é o número de graus de liberdade, e e é uma constante igual a base do sistema de logaritmo natural (aproximadamente 2,71828). Y_0 é definido, de modo que a área sob a curva qui-quadrado seja igual a um.

Na figura abaixo, a curva vermelha mostra a distribuição de valores qui-quadrados calculados de todas amostras possíveis de tamanho 3, onde os graus de liberdade são n – 1 = 3 – 1 = 2. Similarmente, a curva verde mostra a distribuição de amostras de tamanho 5 (graus de liberdade igual a 4); e a curva azul, para amostras de tamanho 11(graus de liberdade igual a 10).

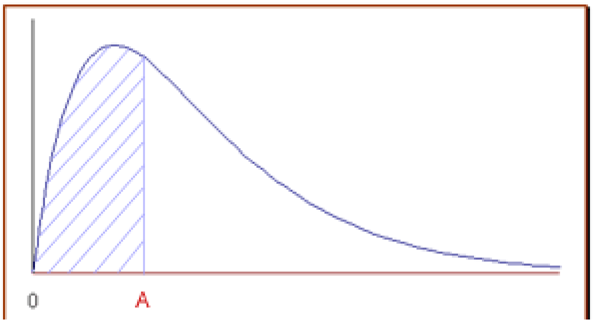


A distribuição qui-quadrado tem as seguintes propriedades:

- A média da distribuição é igual ao número de graus de liberdade: μ = ν.
- A variância é igual a duas vezes o número de graus de liberdade: $\sigma^2 = 2*\nu$
- Quando os graus de liberdade forem maiores que ou iguais a 2, o valor máximo de Y ocorre quando $\chi^2 = \nu - 2$
- Quanto graus de liberdade, mais a curva qui-quadrado se aproxima de uma distribuição normal.

Probabilidade Cumulativa e a Distribuição Qui-Quadrado

A distribuição qui-quadrado é construída de modo que a área total sob a curva seja igual a 1. A área sob a curva entre 0 e um particular valor qui-quadrado é uma probabilidade cumulativa associada com aquele valor qui-quadrado. Por exemplo, na figura abaixo, a área hachuriada representa uma probabilidade cumulativa associada com uma estatística qui-quadrada igual a A; isto é, ela é a probabilidade que o valor de uma estatística qui-quadrado caia entre 0 e A.



Felizmente, não temos que calcular a área sob a curva para encontrar a probabilidade. O modo mais fácil de encontrar a probabilidade cumulativa associada com uma estatística qui-quadrado é usar a planilha Excel.

A Distribuição Qui-Quadrado no Excel

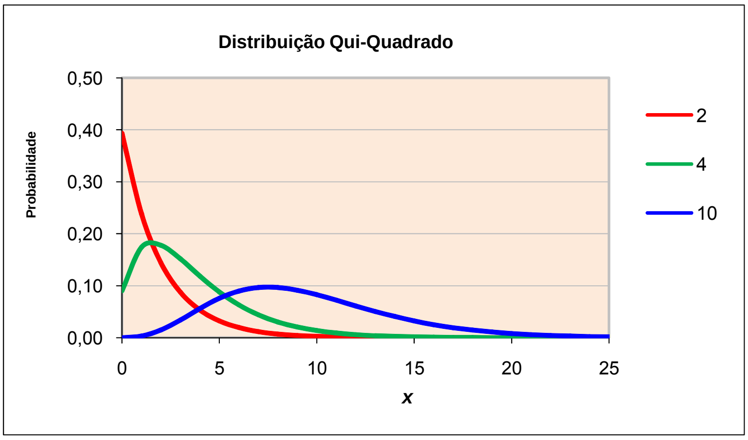
O Excel dispõe das funções estatísticas DIST.QUI, INV.QUI e TESTE.QUI para a distribuição χ^2 cujas sintaxes são as seguintes:

DIST.QUI(x;graus_liberdade)

A função estatística DIST.QUI dá a probabilidade $P(\chi^2 \geq x)$ na cauda superior da distribuição qui-quadrado para os graus_liberdade especificados. Este resultado é o *p-value* na cauda superior da distribuição.

Com a função DIST.QUI foi construída a curva da distribuição qui-quadrado. A fórmula: =DIST.QUI($B5;C$4)-DIST.QUI($B6;C$4) foi, primeiro, registrada na célula C5 e depois copiada no intervalo C5:E30, como mostra a planilha abaixo. Mudando os graus de liberdade do intervalo C4:E4 o modelo construirá outras curvas da distribuição qui-quadrado.

	A	B	C	D	E	F
1	DISTRIBUIÇÃO QUI-QUADRADO					
2						
3			Graus de liberdade			
4		*x*	2	4	10	
5		0	0,3935	0,0902	0,0002	<--=DIST.QUI($B5;E$4)-DIST.QUI($B6;E$4)
6		1	0,2387	0,1740	0,0035	
7		2	0,1447	0,1779	0,0149	
8		3	0,0878	0,1518	0,0341	
9		4	0,0533	0,1187	0,0562	
10		5	0,0323	0,0881	0,0759	
11		6	0,0196	0,0633	0,0898	
12		7	0,0119	0,0443	0,0966	
13		8	0,0072	0,0305	0,0967	
14		9	0,0044	0,0207	0,0916	
15		10	0,0027	0,0139	0,0830	
16		11	0,0016	0,0092	0,0725	
17		12	0,0010	0,0061	0,0614	
18		13	0,0006	0,0040	0,0507	
19		14	0,0004	0,0026	0,0409	
20		15	0,0002	0,0017	0,0324	
21		16	0,0001	0,0011	0,0253	
22		17	0,0001	0,0007	0,0194	
23		18	0,0000	0,0004	0,0147	
24		19	0,0000	0,0003	0,0110	
25		20	0,0000	0,0002	0,0082	
26		21	0,0000	0,0001	0,0060	
27		22	0,0000	0,0001	0,0044	
28		23	0,0000	0,0000	0,0031	
29		24	0,0000	0,0000	0,0023	
30		25	0,0000	0,0000	0,0016	
31						



INV.QUI(probabilidade;graus_liberdade)

A função estatística INV.QUI dá o *valor-crítico* na cauda superior da distribuição qui-quadrado para a probabilidade e os graus_liberdade especificados. A função INV.QUI é a função inversa da DIST.QUI.

TESTE.QUI(intervalo_observado;intervalo_esperado)

A função estatística TESTE.QUI dá a probabilidade $P(\chi^2 \geq x)$ na cauda superior da distribuição qui-quadrado para o intervalo_observado e o intervalo_esperado especificados. Esta função dá o mesmo resultado que a função DIST.QUI.

EXEMPLO 1

A *Nose Battery Company* (NBC) desenvolveu uma nova bateria de telefone celular. Em média, a bateria sobrevive 60 minutos com uma única carga. O desvio padrão é 4 minutos.

Suponha que o departamento de fabricação executa um teste de controle de qualidade. Eles selecionam aleatoriamente 7 baterias. O desvio padrão das baterias selecionadas é 6 minutos. Qual seria a estatística qui-quadrado representada neste teste?

Solução

Sabemos o seguinte:

- O desvio padrão da população é 4 minutos
- O desvio padrão da amostra é 6 minutos
- O número de observações da amostra é 7.

Para calcular a estatística qui-quadrado, liguemos estes dados na equação qui-quadrado, como mostrado abaixo

$$\chi^2 = \frac{[(n-1) * s^2]}{\sigma^2} = \frac{[(7-1) * 6^2]}{4^2} = 13{,}5$$

Onde χ^2 é a estatística qui-quadrado, n é o tamanho da amostra, s é o desvio padrão da amostra, e σ é o desvio padrão da população.

EXEMPLO 2

Vamos revisar o problema apresentado acima. O departamento de fabricação executou um teste de controle qualidade, usando 7 baterias selecionadas aleatoriamente. Nos seus testes, o desvio padrão foi 6 minutos, que igualou à estatística qui-quadrado de 13,5.

Suponha que eles repetiram o teste com uma nova amostra aleatória de 7 baterias. Qual é a probabilidade que o desvio padrão no novo teste será maior que 6 minutos?

Solução

Sabemos o seguinte:

- O desvio padrão da amostra n é igual a 7
- Os graus de liberdade são iguais a n - 1 = 7 - 1 = 6
- A estatística qui-quadrado é igual a 13,5 (ver exemplo 1 acima).

Dados os graus de liberdade, podemos determinar a probabilidade cumulativa que a estatística qui-quadrado caia entre 0 e qualquer valor positivo. Para encontrar a probabilidade cumulativa que uma estatística qui-quadrado caia entre 0 e 13,5, entremos com os graus de liberdade (6) e a estatística qui-quadrado (13,5) na função DIST.QUI da planilha abaixo:

	A	B	C	D
1	**DISTRIBUIÇÃO QUI_QUADRADO**			
2				
3		**tamanho da amostra**	**7,0**	
4		**desvio padrão da amostra**	**6,0**	
5		**desvio padrão da populaçã**	4,0	
6		**qui-quadrado**	13,5	<--=((C3-1)*C4^2)/(C5^2)
7		**graus de liberdade**	6,0	<--C3-1
8		**Probabilidade**	0,04	<--=DIST.QUI(C6;C7)
9		**Probabilidade Cumulativa**	0,96	<--=1-C8
10				

A planilha mostrou que a probabilidade cumulativa é 0,96.

Isto nos diz que a probabilidade que um desvio padrão será menor que ou igual a 6 minutos é 0,96. Isto significa que a probabilidade que o desvio padrão será maior que 6 minutos é 1 - 0,96 = 0,04.

Distribuição F

A estatística F é uma variável aleatória que tem uma distribuição F.

Aqui estão os passos exigidos para se calcular uma **estatística F**:

- Selecione uma *amostra* aleatória de tamanho n_1 de uma população <u>normal</u>, tendo um *desvio padrão*[7] igual a σ_1.
- Selecione uma *amostra* aleatória independente de tamanho n_2 de uma população <u>normal</u>, tendo *desvio padrão* igual a σ_2.
- A **estatística F** é a razão de $\frac{s_1^2}{\sigma_1^2}$ e $\frac{s_2^2}{\sigma_2^2}$.

Assim, para verificar se duas populações independentes têm a <u>mesma variância</u> é utilizada a estatística da relação das <u>variâncias</u> das *amostras* $\frac{s_1^2}{s_2^2}$ retiradas das populações. Se as distribuições das duas populações forem <u>normais</u>, então a relação $\frac{s_1^2}{s_2^2}$ tem **distribuição F**. Sempre que as distribuições das populações forem <u>normais</u>, a distribuição F será utilizada, também, para comparar duas ou mais médias simultaneamente, procedimento denominado *análise da variância*.

As seguintes equações equivalentes são comumente usadas para se calcular uma estatística F:

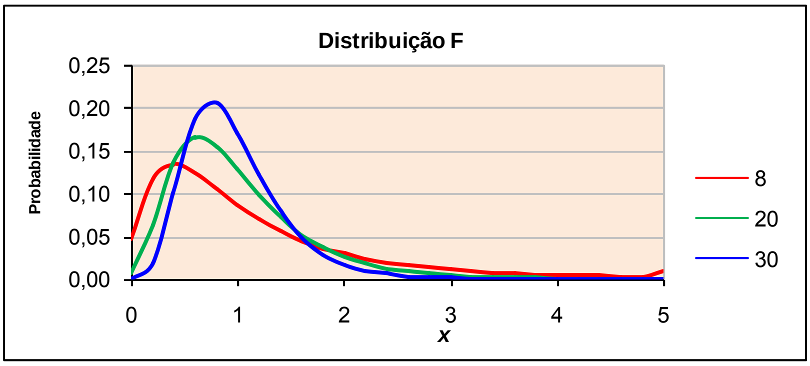
$$F = \frac{\left[\frac{s_1^2}{\sigma_1^2}\right]}{\left[\frac{s_2^2}{\sigma_2^2}\right]}; \qquad F = \frac{[s_1^2 * \sigma_2^2]}{[s_2^2 * \sigma_1^2]}; \qquad F = \frac{\left[\frac{\chi_1^2}{\nu_1}\right]}{\left[\frac{\chi_2^2}{\nu_2}\right]}; \qquad F = \frac{[\chi_1^2 * \nu_2]}{[\chi_2^2 * \nu_1]}$$

Onde σ_1 é o desvio padrão da *população* 1, s_1 é o desvio padrão da *amostra* retirada da população 1, σ_2 é o desvio padrão da *população* 2, s_2 é o desvio padrão da *amostra* retirada da população 2, χ_1^2 é a estatística qui-quadrado para a *amostra* retirada da população 1, ν_1 é o grau de liberdade para χ_1^2, χ_2^2 é a estatística qui-quadrado para a amostra extraída da população 2, e ν_2 são os graus de liberdade para χ_2^2. Note que os graus de liberdade $\nu_1 = n_1 - 1$ e graus de liberdade $\nu_2 = n_2 - 1$.

A distribuição de todos os possíveis valores da *estatística F* é chamada de uma **distribuição F**[8], com $\nu_1 = n_1 - 1$ e $\nu_2 = n_2 - 1$ graus de liberdade.

<u>Características Principais da Distribuição F</u>

- A distribuição F é contínua e sempre positiva com valores no intervalo $(0, +\infty)$
- Há uma família de distribuições F identificadas por dois parâmetros: *graus de liberdade do numerador* ν_1 e *graus de liberdade do denominador* ν_2.
- A distribuição F tem inclinação positiva. A forma final da distribuição depende dos graus de liberdade ν_1 e ν_2, como mostra a figura abaixo



Quando descrevendo uma distribuição F, o número de graus de liberdade associados com o desvio padrão <u>no numerador</u> da estatística F é sempre <u>estabelecido primeiro</u>. Assim, f(5,9) referiremos a uma **distribuição F** com $\nu_1 = 5$ e $\nu_2 = 9$ graus de liberdade. Note que a curva representada por f(5,9) diferirá da curva representada por f(9,5).

A distribuição F tem as seguintes propriedades:

- A média da distribuição é igual a $\nu_1 / (\nu_2 - 2)$.

[7] Desvio padrão da população
[8] Também conhecida como distribuição F de Snedecor

- A variância é igual a $\frac{[\nu_2{}^2 * (\nu_1+2)]}{[\nu_1 * (\nu_2-2) * (\nu_2-4)]}$.

Probabilidade Cumulativa e a Distribuição F

Cada estatística F pode ser associada com uma única probabilidade cumulativa. Esta probabilidade cumulativa representa a probabilidade de que a estatística F seja menor que ou igual a um valor específico.

Os estatísticos usam f_α para representar o valor de uma estatística F tendo uma probabilidade cumulativa de $(1 - \alpha)$. Por exemplo, suponha que estamos interessados na estatística F tendo uma probabilidade cumulativa de 0,95. Referiremos a esta estatística F como $f_{0,05}$, desde que $(1 – 0,95) = 0,05$.

É claro, para encontrar o valor f_α precisaremos saber os graus de liberdade, ν_1 e ν_2. Na notação, os graus de liberdade aparecem entre parênteses como segue: $f_\alpha(\nu_1 , \nu_2)$. Assim, $f_{0,05}(5,7)$ se refere ao valor da estatística F tendo uma probabilidade cumulativa de 0,95, $\nu_1 = 5$ graus de liberdade, e $\nu_2 = 7$ graus de liberdade.

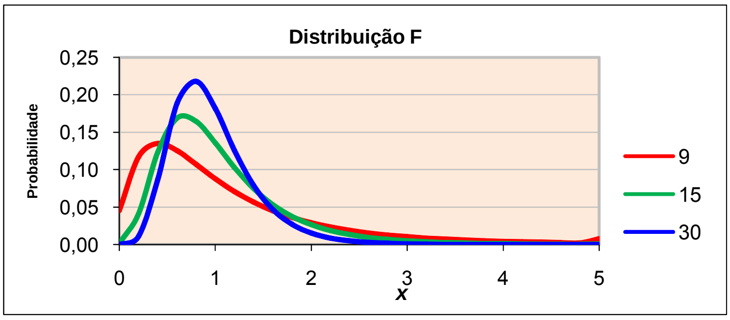
A Distribuição F no Excel

O Excel dispõe das funções estatísticas DISTF e INVF para a **distribuição F** com as seguintes sintaxes:

DISTF(x;gl_numerador;gl_denominador)

A função estatística DISTF dá a probabilidade $P(F \geq x)$ na cauda superior da distribuição F considerando os graus de liberdade do numerador *gl_numerador* e os graus de liberdade do denominador *gl_denominador*. Por exemplo, para x = 14,4 e graus de liberdade gl_numerador = 9 e gl_denominador = 5, com a fórmula: =DISTF(14,4;9;5) obtemos o resultado 0,00451 referente à probabilidade $P(F \geq 14,4)$ na cauda superior da distribuição F.

	A	B	C	D	E
1	**DISTRIBUIÇÃO F**				
2					
3			**Graus de liberdade numerador**		
4			**5**	**15**	**25**
5			**Graus de liberdade denominador**		
6		***x***	**9**	**15**	**30**
7		**0**	0,0455	0,0017	0,0001
8		**0,2**	0,1172	0,0413	0,0110
9		**0,4**	0,1351	0,1235	0,0870
10		**0,6**	0,1255	0,1690	0,1887
11		**0,8**	0,1071	0,1644	0,2180
12		**1**	0,0878	0,1357	0,1813
13		**1,2**	0,0708	0,1030	0,1259
14		**1,4**	0,0567	0,0749	0,0790
15		**1,6**	0,0453	0,0533	0,0468
16		**1,8**	0,0363	0,0376	0,0269
17		**2**	0,0291	0,0265	0,0153
18		**2,2**	0,0235	0,0187	0,0086
19		**2,4**	0,0191	0,0133	0,0049
20		**2,6**	0,0156	0,0096	0,0028
21		**2,8**	0,0128	0,0069	0,0016
22		**3**	0,0106	0,0050	0,0009
23		**3,2**	0,0088	0,0037	0,0005
24		**3,4**	0,0073	0,0027	0,0003
25		**3,6**	0,0061	0,0020	0,0002
26		**3,8**	0,0052	0,0015	0,0001
27		**4**	0,0044	0,0012	0,0001
28		**4,2**	0,0037	0,0009	0,0000
29		**4,4**	0,0032	0,0007	0,0000
30		**4,6**	0,0027	0,0005	0,0000
31		**4,8**	0,0023	0,0004	0,0000
32		**5**	0,0078	0,0011	0,0000



INVF(probabilidade;gl_numerador;gl_denominador)

A função estatística INVF dá o *F crítico* (F_c) da **distribuição F** quando conhecida a probabilidade na cauda superior da distribuição F, e os graus de liberdade do numerador e do denominador. A função INVF é a função inversa da DISTF. Por exemplo, para a probabilidade 0,00451, gl_numerador = 9 e gl_denominador = 5, a fórmula: INVF(0,00451;9;5) dá o *F* *crítico* como $F_c = 14{,}40$. Como o cálculo de F_c é um procedimento iterativo, se depois de realizar 100 iterações não for alcançado o resultado com um erro de $\pm 3 \times 10^{-7}$, a função INVF apresentará o resultado #N/A.

EXEMPLO 1

Suponha que você selecionou aleatoriamente 7 mulheres de uma população de mulheres, e 12 homens de uma população de homens. A tabela abaixo mostra os desvios padrões de cada amostra e de cada população.

População	Desvio Padrão da População	Desvio Padrão da Amostra
Mulheres	30	35
Homens	50	45

Calcule a **estatística F.** ν_1 e ν_2

Solução

A estatística F pode ser calculada a partir dos desvios padrões da população e da amostra, usando a seguinte equação:

$$F = \frac{\left[\frac{s_1^2}{\sigma_1^2}\right]}{\left[\frac{s_2^2}{\sigma_2^2}\right]}$$

Onde σ_1 é o desvio padrão da ***população*** 1, s_1 é o desvio padrão da ***amostra*** retirada da população 1, σ_2 é o desvio padrão da ***população*** 2, s_2 é o desvio padrão da ***amostra*** retirada da população 2,

Como você pode ver da equação, existem realmente duas maneiras de se calcular uma estatística F desses dados. Se os dados das mulheres aparecem no numerador, podemos calcular uma estatística F como segue:

$$F = \frac{\left[\frac{35^2}{30^2}\right]}{\left[\frac{45^2}{50^2}\right]} = \frac{\left[\frac{1.225}{900}\right]}{\left[\frac{2025}{2500}\right]} = \frac{1.361}{0{,}81} = 1{,}68$$

Para este cálculo, os graus de liberdade do numerador são $\nu_1 = 7 - 1 = 6$; e os do denominador $\nu_2 = 12 - 1 = 11$.

Na planilha teremos:

	G	H	I	J	K
1	DesvPad Pop 1	30			
2	DesvPad Pop2	50			
3	DesvPad Amostra 1	35			
4	DesvPad Amostra 2	45			
5	Estatística F	1,680384	<--=((H3^2/H1^2)/(H4^2/H2^2))		

Por outro lado, se os dados dos homens aparecem no numerador, calculamos a estatística F como segue:

$$F = \frac{\left[\frac{45^2}{50^2}\right]}{\left[\frac{35^2}{30^2}\right]} = \frac{\left[\frac{2.025}{2.500}\right]}{\left[\frac{1.225}{900}\right]} = \frac{0{,}81}{1{,}361} = 0{,}595$$

Para este cálculo, os graus de liberdade do numerador são $\nu_1 = 12 - 1 = 11$; e os do denominador $\nu_2 = 7 - 1 = 6$.

Na planilha teremos:

	G	H	I	J	K
1	DesvPad Pop 1	50			
2	DesvPad Pop2	30			
3	DesvPad Amostra 1	45			
4	DesvPad Amostra 2	35			
5	Estatística F	0,595102	<--=((H3^2/H1^2)/(H4^2/H2^2))		

EXEMPLO 2

Encontre a *probabilidade cumulativa associada* com cada uma das **estatísticas F** do Exemplo 1, acima.

Solução

Para resolver este problema, precisamos encontrar os graus de liberdade de cada amostra. Depois então, usaremos a planilha Excel que realiza os cálculos para encontrar as probabilidades.

- Os graus de liberdade da amostra de mulheres são iguais a n – 1 = 7 – 1 = 6.
- Os graus de liberdade da amostra de homens são iguais a n – 1 = 12 – 1 = 11.

Portanto, quando os dados das mulheres aparecerem no numerador, os graus de liberdade do numerador ν_1 são iguais a 6; e os do denominador ν_2 = 11. E, baseado nos cálculos mostrados no exemplo anterior, a estatística F é igual a 1,680384. Levando estes valores à planilha encontramos que a probabilidade cumulativa é **0,7844**.

	N	O	P	Q	R
1	Função DISTF			Função INVF	
2					
3	DesvPad Pop 1	30		*Probabilidade*	0,2156
4	DesvPad Pop2	50		*gl numerador*	6
5	DesvPad Amostra 1	35		*gl denominador*	11
6	DesvPad Amostra 2	45		*F crítico(0,2156;6;11)*	1,680
7	Estatística F	1,680384	<--=((O5^2/O3^2)/(O6^2/O4^2))		
8	*x*	1,680384	<--=O7		
9	*gl numerador*	6			
10	*gl denominador*	11			
11	*P(F >=1,6803840877915)*	0,2156	<--=DISTF(O8;O9;O10)		
12	*P(F<1,6803840877915)*	0,7844	<--=1-O11		
13					

Por outro lado, quando os dados dos homens aparecerem no numerador, os graus de liberdade do numerador ν_1 são iguais a 11; e os do denominador ν_2 = 6. E, baseado nos cálculos mostrados no exemplo anterior, a estatística F é igual a 0,595102. Levando estes valores à planilha encontramos que a probabilidade cumulativa é **0,2156**.

	N	O	P	Q	R
1	Função DISTF			Função INVF	
2					
3	DesvPad Pop 1	50		*Probabilidade*	0,2156
4	DesvPad Pop2	30		*gl numerador*	6
5	DesvPad Amostra 1	45		*gl denominador*	11
6	DesvPad Amostra 2	35		*F crítico(0,2156;6;11)*	1,680
7	Estatística F	0,595102	<--=((O5^2/O3^2)/(O6^2/O4^2))		
8	*x*	0,595102	<--=O7		
9	*gl numerador*	11			
10	*gl denominador*	6			
11	*P(F >=0,595102040816327)*	0,7844	<--=DISTF(O8;O9;O10)		
12	*P(F<0,595102040816327)*	0,2156	<--=1-O11		

Distribuição de Weibull

A distribuição de probabilidade Weibull é uma distribuição de probabilidade contínua amplamente utilizada na análise de dados de vida de equipamentos devido a sua flexibilidade – ela pode imitar outras distribuições de probabilidade, como a distribuição exponencial e a distribuição normal, dependendo do valor de seus parâmetros.

O seu nome se deve ao seu inventor, Waloddi Weibull, e é usada extensivamente em engenharia de confiabilidade e no cálculo do tempo médio de falha para determinado dispositivo.

As principais vantagens da utilização da distribuição de Weibull para análise da sobrevivência é que através da estimativa de apenas dois parâmetros (alfa e beta) são obtidas informações tanto de longevidade média quanto do tipo de curva de sobrevivência. Outra vantagem é que as observações não necessitam ser realizadas a intervalos constantes, como, por exemplo, com as tabelas de esperança de vida.

A fdp da distribuição Weibull é descrita pela Equação:

$$f(x;\beta;\alpha) = \frac{\beta}{\alpha^{\beta}} x^{\beta-1}\, e^{-\left(\frac{x}{\alpha}\right)^{\beta}} \qquad \beta>0 \text{ e } \alpha>0$$

Onde:

- β é o parâmetro de forma (shape);
- α é o parâmetro de escala (scale);

A fda[9] da distribuição Weibull é descrita pela Equação:

$$F(x;\beta;\alpha) = 1 - e^{-\left(\frac{x}{\alpha}\right)^{\beta}}$$

O parâmetro β influencia na fdp da distribuição Weibull da seguinte forma:

- Para $0 < \beta \leq 1$:
  - $f(t) \to \infty$ quando $t \to 0$;
  - $f(t) \to 0$ quando $t \to \infty$.
- Para $\beta > 1$:
  - $f(t) = 0$ quando $t = 0$;
  - $f(t)$ cresce quanto $t \to \tilde{t}$ (moda) e decresce logo após.

O Fator de Forma β (indica a forma da curva e a característica das falhas).

- "$\beta < 1$" mortalidade infantil
- "$\beta = 1$" falhas aleatórias (função exponencial negativa)
- "$\beta > 1$" falhas por desgaste

Observações relativas ao Fator de Forma "β":

A escolha apropriada de "β" e "α" na Distribuição de Weibull pode ser usada para representar uma larga faixa de distribuições, incluindo tanto distribuições randômicas (exponencial negativa) quanto às distribuições aproximadamente normais. Embora a experiência tenha mostrado que a distribuição de Weibull possa ser usada para representar a grande maioria de modelos de falha, é essencial notar que é uma função semi-empírica, e pode não ser capaz de representar algumas distribuições particulares encontradas na prática.

Com relação ao Fator de Forma "β", temos que:

- Se "$\beta = 1$" (taxa de falha constante), pode ser uma indicação que modos de falhas múltiplos estão presentes ou que os dados coletados dos tempos para falhar são suspeitos. Este é freqüentemente o caso dos sistemas os quais diferentes componentes têm diferentes idades, e o tempo individual de operação dos componentes não

[9] Função distribuição acumulada

estão disponíveis. Uma taxa de falhas constante pode também indicar que as falhas são provocadas por agentes externos, tais como: uso inadequado do equipamento ou técnicas inadequadas de manutenção.

- O modo de falhas por desgaste é caracterizado por "$\beta > 1$", mas pode ocorrer situações as quais as falhas por desgaste ocorram depois de um tempo finito livre de falhas, e um valor de "$\beta = 1$" é obtido. Isto pode ocorrer quando uma amostragem contém uma proporção de itens imperfeitos, acarretando falhas antes de um tempo finito livre de falhas. Os parâmetros da Distribuição de Weibull dos modos de falhas por desgaste podem ser deduzidos se forem eliminados os itens imperfeitos e analisados os seus dados separadamente.

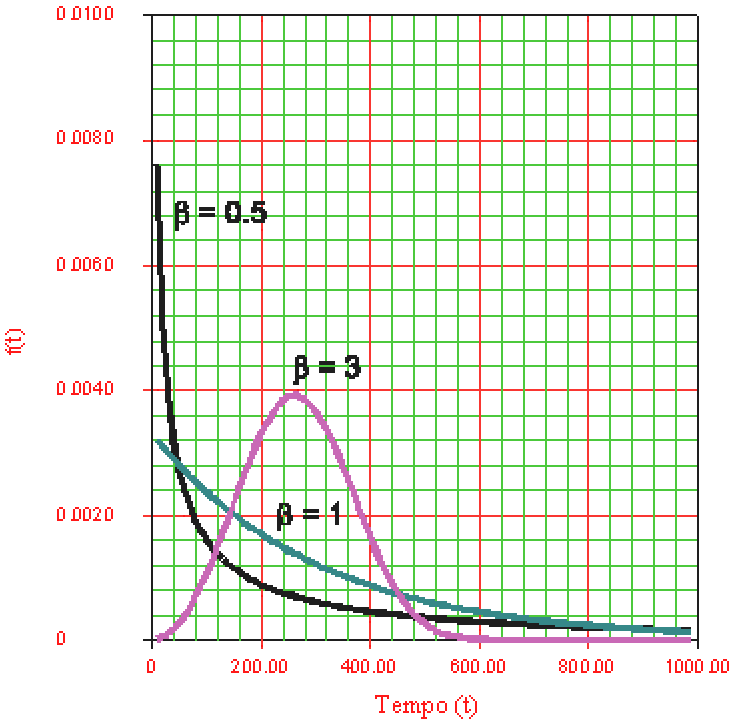


Figura A - Efeito do parâmetro β na fdp. Retirado de [Life Data Analysis[10]].

O parâmetro α influencia na fdp da distribuição Weibull da seguinte forma (Figura 2-2):

- Se α cresce enquanto β é constante, a fdp se estica para a direita e sua altura diminui;
- Se α decresce enquanto β é constante, a fdp se encolhe para a esquerda e sua altura aumenta.

[10] http://www.weibull.com/lifedatawebcontents.htm

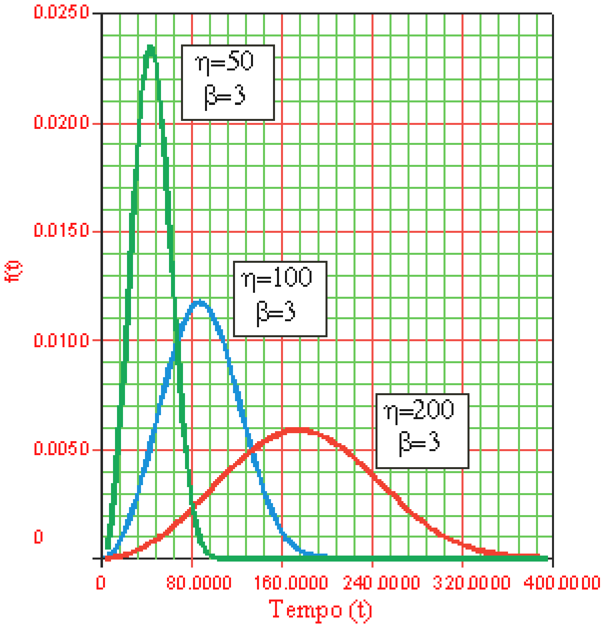


Figura B - Efeito do parâmetro α na fdp. Retirado de [Life Data Analysis].

A distribuição exponencial (usada para estudar tempo de espera) é um caso especial da distribuição Weibull com alfa = 1, média = beta e lâmbda(a taxa de risco)=1/beta. Outro caso especial da distribuição Weibull é a distribuição Rayleigh (usada para estudar o espalhamento de radiação, velocidade de ventos ou fazer certas transformações). Para a distribuição Rayleigh alfa é fixado em 2.

Na prática a distribuição Weibull é usada para descrever dois grupos de fenômenos. O tempo de vida de objetos é frequentemente usado em controle de qualidade. Um fabricante fornece os parâmetros Weibull para um produto e o usuário pode calcular a probabilidade que uma parte falhe após um, dois, três ou mais anos. O programa distribuição Weibull permite-lhe fazer estes cálculos com base nos parâmetros já conhecidos. Por exemplo, se você quiser saber a proporção que falha após um ou mais anos, entre com o valor um na caixa 'x' e leia o valor da probabilidade acumulada. Se você quiser saber o momento no tempo em que as partes foram divididas você fracassará, entre com o valor 0.5 caixa '%' e leia o valor de 'x'.

A descrição da velocidade dos ventos é um exemplo do uso da distribuição Weibull para descrever fenômenos naturais. Cada parte do planeta tem os seus próprios parâmetros para uma distribuição Weibull para descrever o modelo da velocidade dos ventos naquele lugar. Com base nisto você pode calcular o número de dias por ano, ou horas por dia, com velocidade dos ventos acima de certa força, ou a média da velocidade dos ventos, ou a mediana da velocidade dos ventos, dividir em dois os dias do ano e ter uma velocidade dos ventos abaixo da força média, metade dos dias acima. A distribuição Weibull muito prática nesta área porque a distribuição não permite valores negativos e é fácil de considerar apropriadamente o fato que na maioria dos dias existirão um pouco de vento, em alguns dias uma porção e você tem aqueles dias que existem muito mais vento.

A distribuição-log Weibull se concentra no log de uma variável distribuída por Weibull. Ela dá o limite da distribuição para os menores e os maiores valores nas amostras extraídas de uma variedade de distribuições. A distribuição é usada para descrever condições extremas, tais como rajada de vento extrema, energia extrema liberada durante terremotos, ou stress extremos para os quais os componentes estão sujeitos. Algumas vezes a distribuição é usada como uma alternativa à distribuição normal no caso de dados assimétricos. Outros nomes para a log-Weibull são "distribuição Fisher-Tippett" ou "distribuição *extreme value*". Embora a distribuição mais usada seja a distribuição *extreme value* existem outras distribuições de valores extremos descrevendo a distribuição limite para os menore e os maiores valores extraídos de uma particular distribuição. A distribuição de Gumbel é um caso especial de log distribuição Weibull. Para a distribuição Gumbel alfa=0 e beta=1.

Existem vários pacotes estatísticos para estimar os parâmetros Weibull para um conjunto de dados. Não existe portanto muitos pacotes para a log-Weibull. Você terá de procurar por eles na Internet. Infelizmente estes pacotes tendem a ser caros.

Weibull no Excel

O Excel possui a função WEIBULL com a seguinte sintaxe:

WEIBULL(x;beta;alfa;cumulativo)

X é o valor no qual se avalia a função.

Alfa é um parâmetro da distribuição

Beta é um parâmetro da distribuição

Cumulativo determina a forma da função

Quando alfa = 1, a WEIBULL retornará a distribuição exponencial com : $\lambda = 1/\beta$.

Por exemplo: =WEIBULL(105;20;100;FALSO) dá 0,035589

=WEIBULL(105;20;100;VERDADEIRO) dá 0,929581

Você poderia também usar o procedimento que desenvolvemos em *Javascript* para a realização deste cálculo. Assim

O link[11] é:

http://www.bertolo.pro.br/FinEst/Estatistica/DistribuicaoProbabilidades/binomial.htm

[11] Outras distribuições poderão ser calculadas neste site: http://www.bertolo.pro.br/FinEst/Estatistica/index.html