Métodos Estatísticos de Apoio à Decisão

Aula 5

Mônica Barros, D.Sc.

Agosto de 2008

Programa do Curso

Disciplina	Métodos Estatísticos de Apoio à Decisão - BI MASTER 2008		
Responsável	Mônica Barros		
Ferramentas	Excel, @Risk		

Aula	Tipo (T-P-C)	Tema	Descrição
1	T, P	Estatística Descritiva	Gráficos, tabelas e medidas numéricas
2	T	Probabilidade: Definições básicas	Definições básicas: probabilidade, espaço amostral, eventos, propriedades das probabilidades, Probabilidade Condicional, Independência;Teorema de Bayes
3	T	Probabilidade: Definições básicas	Variáveis Aleatórias Contínuas e Discretas , Função de Probabilidade, Função Densidade, Função de Distribuição, Momentos de uma v.a., Média, Variância e Desvio Padrão
4	T, P	Probabilidade: Definições básicas	Variáveis Discretas: Bernoulli, Binomial, Hipergeométrica, Geométrica, Binomial Negativa, Poisson
5	T, P	Probabilidade: v.a. Contínuas	Variáveis Contínuas: Uniforme, Exponencial, Normal
6	P	Pratica 1	Aula de exercícios - As funções do Excel para cálculo de probabilidades para v.a. Contínuas e discretas
7	T, C	Probabilidade: v.a. Contínuas E CASE 1: Simulação - soma de v.a. e o teorema central do limite CASE 2: Otimização de um portfolio simulado - propriedades da média e variância e o uso do Solver	O teorema central do limite e a importância da distribuição Normal.O teorema central do limite na prática - soma de variáveis aleatórias e a convergência para a Normal. Distribuição da soma de v.a. e da média amostral. Propriedades da média e variância de combinações lineares de v.a. - o efeito da correlação. O uso do Solver do Excel
8	T, P	Distribuições Amostrais	Amostra aleatória simples, distribuição da média amostral, distribuição de p^
9	T, P	Estatística - estimação pontual	Estimação da média da população com sigma conhecido e desconhecido e para proporções
10	T/P	Estatística - estimação por intervalos	Intervalos de confiança para amostras Normais e proporção Binomial - Exercícios - intervalos de confiança empregando o Excel
11	T/P	Estatística - testes de hipóteses	Teste de hipótese para amostrais normais e Exercícios

Aula 5

- Variáveis Contínuas
  - Uniforme
  - Exponencial
  - Normal
  - Lognormal

Distribuição Uniforme

- A probabilidade de ocorrência em dois intervalos quaisquer de mesmo tamanho é a mesma – a função de densidade de probabilidade é uma reta paralela ao eixo horizontal.

- Se considerarmos os limites de ocorrência de x como sendo <u>a</u> e <u>b</u> (a < b) devemos ter necessariamente f(x) = 1/(b – a) para que a integral da densidade seja 1.

Distribuição Uniforme

- Se X ~ Unif(a,b) então sua densidade é:

$$f(x)=\begin{cases}\dfrac{1}{b-a} & \text{se } x \in (a,b)\\ 0 & \text{se } x \notin (a,b)\end{cases}$$

- A função de distribuição é dada por:

$$F(x)=\Pr(X \le x)=\begin{cases}0 & \text{se } x \le a\\ \dfrac{x-a}{b-a} & \text{se } x \in (a,b)\\ 1 & \text{se } x \ge a\end{cases}$$

- Note que a função de distribuição é *linear* no intervalo (a,b).

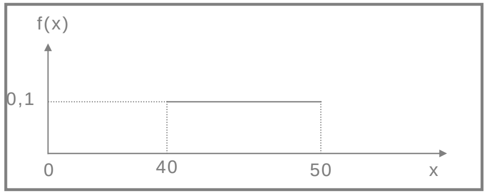
Distribuição Uniforme

- Média e Variância da distribuição Uniforme
- Se X ~ Unif(a,b) então:

$$E(X)=\frac{a+b}{2},\ VAR(X)=\frac{(b-a)^2}{12}$$

Distribuição Uniforme

- Exemplo
- Um vôo da ponte aérea RJ-SP leva entre 40 e 50 minutos, com igual probabilidade de ocorrência dentro desse intervalo
  - A distribuição é Uniforme no intervalo (40, 50)
  - f(x) = 1/(50 – 40) para x no intervalo (40,50) e zero fora desse intervalo



Distribuição Uniforme

- Qual a probabilidade de um vôo durar mais de 48 minutos?

$$\Pr(X>48)=\int_{48}^{50}\frac{1}{50-40}dx=\frac{2}{10}$$

- Qual a probabilidade de um vôo durar entre 43 e 45 minutos?

$$\Pr(43<X<45)=\int_{43}^{45}\frac{1}{50-40}dx=\frac{2}{10}$$

- Uma característica importante da densidade Uniforme é: dois subintervalos de comprimento l que estão totalmente "dentro" de (a, b) têm a mesma probabilidade. Isso não ocorre em geral, no caso de outras densidades.

Distribuição Uniforme

- Exemplo

  O peso mínimo de um pacote de 1Kg de café é de 0,98Kg. O fabricante garante que a distribuição de pesos é uniforme e que a função de densidade de probabilidade, f(x), é igual a 9,75. Se o fabricante disse a verdade, qual é o peso máximo que um pacote de café pode ter?

- Solução

  Seja b o peso máximo. Se a distribuição é uniforme, a área sob f(x) no intervalo de validade de x deve ser igual a 1.

  A área é dada por f(x).(b – a), onde f(x) = 9,75 e a = 0,98

  Logo, b = a + 1/f(x) = 0,98 + 1/9,75 = 1,0826

Distribuição Uniforme

- Exemplo (para casa)
- O retorno de uma aplicação financeira de risco num intervalo de uma semana é uma variável com distribuição Uniforme no intervalo –2% a 1.8%. Calcule:

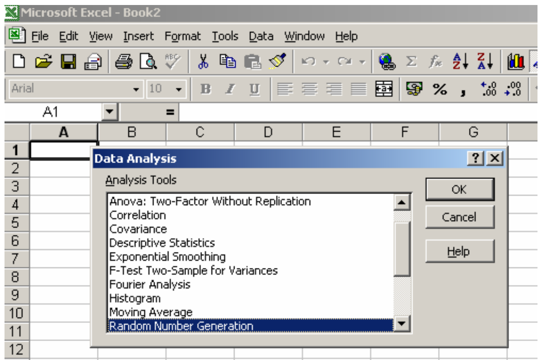
  A probabilidade do retorno do investimento nesta semana ser positivo.

  A probabilidade do retorno estar entre –1% e +1%.

  A probabilidade do retorno exceder 0.5%.

Distribuição Uniforme

- Geração de v.a. Uniformes no Excel
- (É necessária a instalação prévia do suplemento “Análise de Dados”)

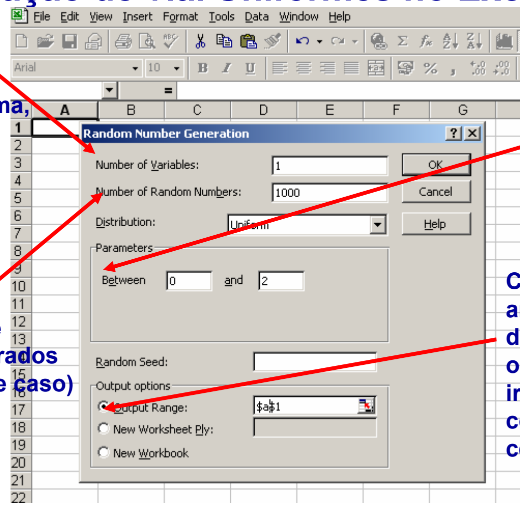


Distribuição Uniforme

- Geração de v.a. Uniformes no Excel

Número de variáveis geradas (uma, neste caso)

Número de valores gerados (1000 neste caso)



Intervalo de definição, neste caso, densidade Unif(0,2)

Célula inicial de armazenamento dos dados – neste caso os números gerados irão preencher a coluna A, a partir da célula A1

Densidade Exponencial

- Serve para:
  - Modelar tempos de duração de equipamentos;
  - Modelar tempos entre ocorrências, por exemplo, o tempo entre chegadas de carros num pedágio, entre a chegada de pessoas num caixa de banco;
- Densidade

$$f(x) = \lambda.\exp(-\lambda.x) \quad \text{onde } \lambda > 0 \quad \text{e} \quad x \geq 0$$

- Função de Distribuição

$$F(x) = \Pr(X \leq x) = \int_0^x \lambda.\exp(-\lambda.u)du = -e^{-\lambda.u}\Big|_0^x = 1 - e^{-\lambda.x}$$

Distribuição Exponencial

- Média e Variância
  - Se X é Exponencial com parâmetro λ, então:

$$E(X) = 1/\lambda$$

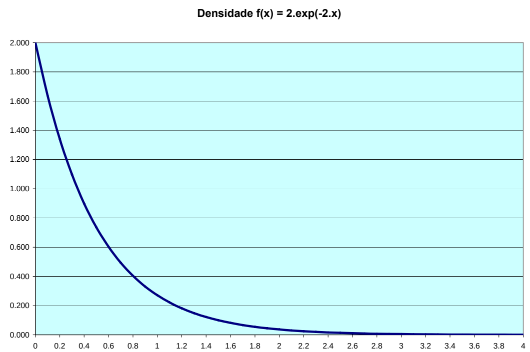
$$VAR(X) = 1/\lambda^2$$

- Falta de Memória
- A distribuição Exponencial "não tem memória". O que isso quer dizer? Esta propriedade indica que a vida restante de um equipamento não depende da idade atual deste equipamento. Ou seja, um componente usado é tão bom quanto um novo (em termos da sua durabilidade).

Densidade Exponencial

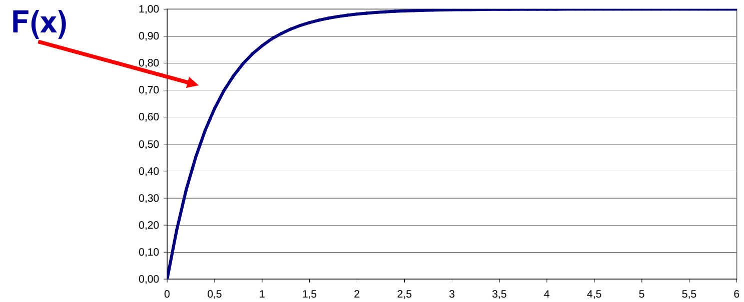
- Gráfico – densidade Exponencial com $\lambda = 2$

$$f(x) = 2.\exp(-2.x) \quad \text{onde } x \geq 0$$



Densidade Exponencial

- O gráfico a seguir apresenta a função de distribuição de uma v.a. Exponencial com parâmetro $\lambda = 2$, isto é, a função de distribuição associada à densidade da página anterior.



Densidade Exponencial

- Exemplo
- O tempo entre as chegadas de táxi num cruzamento é uma variável Exponencial com λ = 1/10 chegadas por minutos. Calcule:

  a) A probabilidade de alguém ter que esperar mais de 60 minutos por um táxi.

  b) A probabilidade de um táxi demorar menos de 10 minutos para passar.

Densidade Exponencial

- Solução

  Seja T o tempo entre chegadas de um táxi, isto é, o tempo que você terá que esperar por um táxi nesta esquina.

  T é uma variável Exponencial com λ = 1/10.

  Para uma variável Exponencial, a função de distribuição é $F(t) = Pr(T \leq t) = 1 - \exp(-\lambda.t)$ e também $Pr(T > t) = 1 - F(t) = \exp(-\lambda.t)$. Logo:

a) $Pr(T > 60) = \exp(-60/10) = \exp(-6) = 0.0025$

b) $Pr(T < 10) = 1 - \exp(-10/10) = 1 - \exp(-1) = 0.6321$

Densidade Exponencial

- Exemplo (para casa)
- O tempo até a ocorrência de um defeito (isto é, o tempo de duração) numa TV é uma variável Exponencial com parâmetro λ = 1/3 anos.
- Calcule a probabilidade de uma TV "pifar" nos primeiros 2 anos de uso.
- Calcule a probabilidade de uma TV durar mais de 5 anos.
- Calcule a probabilidade de uma TV durar entre 3 e 5 anos.

Densidade Exponencial

- Exemplo - Simulação
- A maioria das linguagens de programação tem um gerador de variáveis Uniforme (0,1) "embutido".
- Mas, é conveniente ser capaz de gerar variáveis com outras densidades.
- Pode-se mostrar (e faremos isso eventualmente) que, se U ~ Unif(0,1) então:

$$Y = \frac{-1}{\lambda}.\log(U)$$

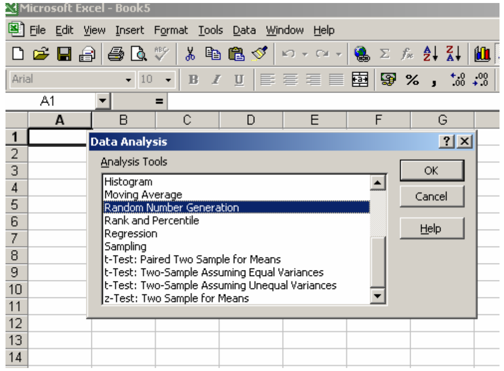
- tem densidade Exponencial com parâmetro λ.

Densidade Exponencial

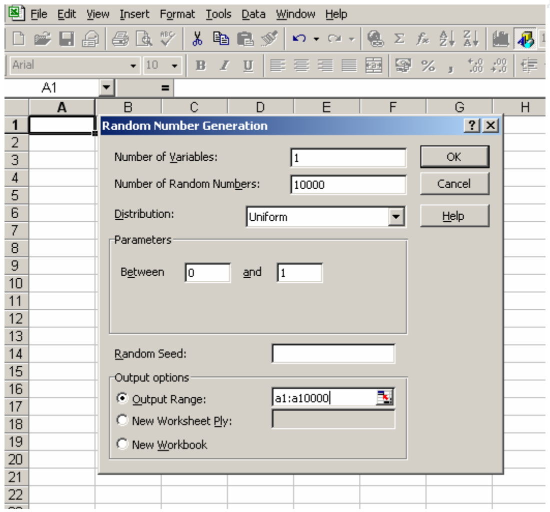
- O próximo exemplo apresenta a geração de 10000 variáveis Exponenciais com parâmetro 1 a partir de uma amostra do mesmo tamanho da Uniforme(0,1).
- Neste exemplo usamos o suplemento "Análise de dados" do Excel, que permite a geração de v.a. e a construção dos histogramas indicados.

Variável Exponencial - simulação

- Suponha que geramos uma amostra aleatória de 10000 observações da densidade Unif(0,1) no Excel, como mostrado nas próximas figuras.

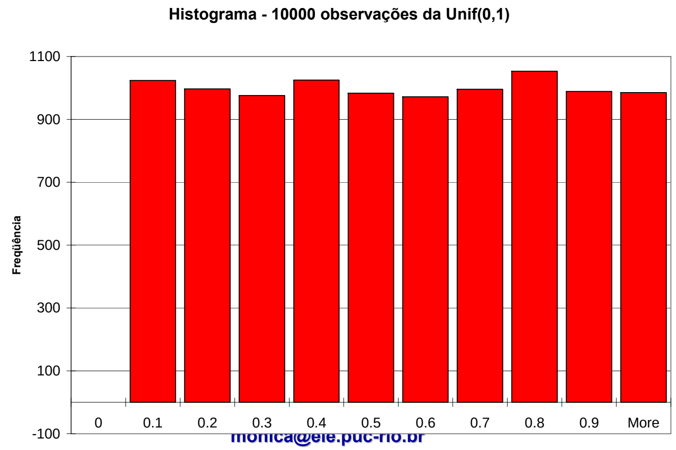


Variável Exponencial - simulação



Variável Exponencial - simulação

- O histograma das 10000 observações geradas é:

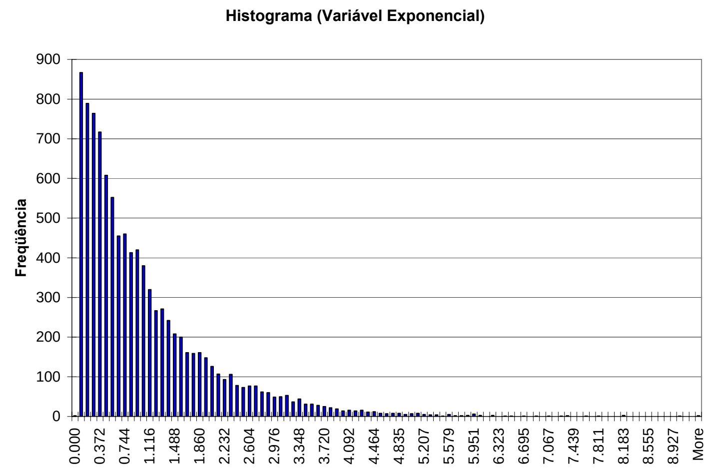


Variável Exponencial - simulação

- Agora criamos uma nova coluna de 10000 observações usando a transformação Y = - log (U) onde U é um valor gerado da distribuição Unif(0,1).

- O histograma da nova amostra deve ter um comportamento decrescente, que se "pareça" com uma densidade Exponencial com média 1. Este histograma é mostrado na próxima figura.

Variável Exponencial - simulação



Variável Exponencial - simulação

- Nota:
- O Excel não tem um gerador de variáveis Exponenciais. O procedimento que você deve usar para simulá-las é apenas uma extensão do método mostrado neste exemplo.
- Para gerar uma variável X com densidade:

$$f(x) = \lambda.\exp(-\lambda.x) \quad \text{onde } \lambda > 0 \ \text{ e } \ x \geq 0$$

- Faça X = (-1/λ). Log(U) onde U é uma v.a. Uniforme(0,1).

Distribuição Normal

- A distribuição Normal é talvez a mais importante das distribuições de probabilidade.

- Muitos fenômenos físicos ou econômicos são freqüentemente modelados pela distribuição Normal.

- É utilizada para descrever inúmeras aplicações práticas:
  - Altura e peso de pessoas e objetos
  - Nível de chuvas
  - Altura de árvores em uma floresta

Distribuição Normal

- A distribuição Normal tem a forma de um sino, e possui **dois parâmetros, μ e σ^2** .
- A distribuição Normal é também chamada de **Gaussiana** em homenagem ao matemático Carl Friederich Gauss (1777 - 1855).
- A distribuição Normal também funciona como uma **boa aproximação para outras densidades**. Por exemplo, sob algumas condições pode-se provar que a densidade Binomial pode ser aproximada pela Normal.

Distribuição Normal

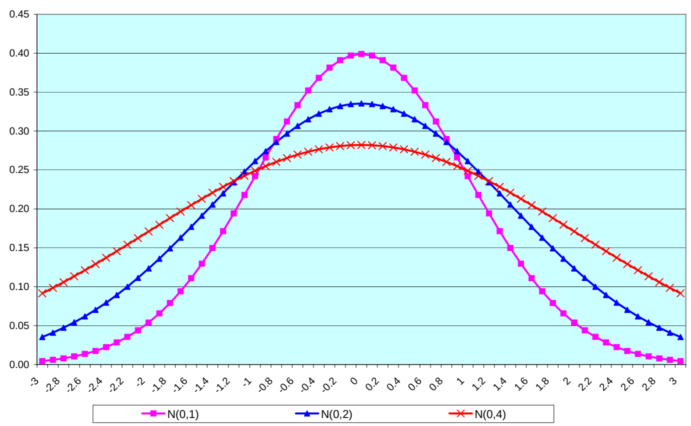
- Densidade Normal com média μ e variância σ^2

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} . e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{onde } \sigma^2 > 0 \; e \; \mu \in R$$

- Notação: X ~ N(μ , σ^2)
- A densidade é simétrica em torno de μ, e quanto maior o valor da variância σ^2, mais "espalhada" é a distribuição.

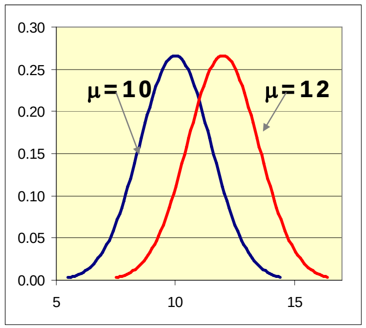
Distribuição Normal

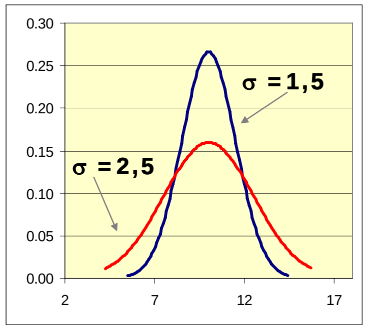
Densidades Normais com média zero e variâncias 1, 2 e 4



Distribuição Normal

- A distribuição normal é completamente caracterizada por sua média μ e seu desvio-padrão σ
- A média define o deslocamento horizontal da curva, enquanto o desvio-padrão define o seu achatamento





Distribuição Normal

- **Propriedades**
  1) f(x) como definida integra a 1.
  2) f(x) > 0 sempre.
  3) Os limites de f(x) quando x tende a + ∞ e - ∞ são iguais a zero.
  4) A densidade $N(\mu, \sigma^2)$ é *simétrica em torno de* μ, ou seja:

     $f(\mu + x) = f(\mu - x)$
  5) O valor máximo de f(x) ocorre em $x = \mu$
  6) Os pontos de inflexão de f(x) são $x = \mu + \sigma$ e $x = \mu - \sigma$.

Distribuição Normal

- **Média, Variância e função de distribuição**
- **Se $X \sim N(\mu, \sigma^2)$ então:**

  $E(X) = \mu$,

  $VAR(X) = \sigma^2$

- **A sua função de distribuição é:**

$$F(x) = \Pr(X \le x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-(u-\mu)^2}{2\sigma^2}\right\} du$$

Não é possível resolver analiticamente esta integral – precisamos de uma tabela!

Distribuição Normal

- **Tabela: será feita para a distribuição N(0,1)**
- É possível transformar uma variável $N(\mu,\sigma^2)$ numa N(0,1) sem grandes dificuldades, e então podemos **tabelar os valores da função de distribuição de uma N(0,1)**, e esta tabela pode ser usada para encontrar probabilidades envolvendo qualquer variável aleatória Normal.

Distribuição Normal

- **Problema:**

  Não é possível criar uma tabela para cada uma das (infinitas) densidades Normais existentes.
- **Solução:**

  Trabalha-se com a densidade Normal com média 0 e variância 1, e converte-se todas as outras Normais para esta, chamada de Normal padrão ou Normal standard.

  A maioria dos livros de estatística fornece tabelas de probabilidade para a distribuição normal padronizada.

Distribuição Normal

- **Transformação numa N(0,1)**
- **Se X ~ N(μ , σ^2) então Z = (X – μ)/σ é uma variável Normal com média 0 e variância 1.**

- **Logo, para transformar uma variável aleatória Normal com quaisquer parâmetros numa Normal (0,1) você deve:**
  - 1- Subtrair a média
  - 2- Dividir o resultado por σ, o desvio padrão

A variável aleatória resultante deste procedimento é uma N(0,1).

Distribuição Normal

- **Se X pertence a uma distribuição normal com média μ e desvio-padrão σ, seu valor normalizado é dado por:**

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

A variável Z é Normal com média 0 e variância 1

- **Existem dois tipos de tabela, que fornecem basicamente a mesma coisa:**
  - $Pr(0 \leq Z \leq z_0)$, ou seja, a probabilidade do lado direito da curva normal a partir da média até o valor z_0
  - $\Phi(z_0) = Pr (Z \leq z_0) = 0.5 + Pr (0 \leq Z \leq z_0)$ (por que?)

- **Iremos trabalhar com a tabela da função de distribuição, isto é: $\Phi(z_0)$**

Distribuição Normal

- **Toda variável Normal pode ser transformada numa Normal com média 0 e variância 1.**

- **Logo, só existe a necessidade de criar uma única tabela para a função de distribuição acumulada.**

- **Se X é N(μ , σ^2) . Então a variável Z = (X - μ) /σ tem distribuição Normal com média zero e variância um, isto é, Z é N(0,1).**

Distribuição Normal

- **Cálculo de probabilidades**

Se X é uma variável Normal com média μ e desvio padrão σ então:

$$\Pr(a \leq X \leq b) = \Pr\left(\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) = \Pr\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right) =$$

$$= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

- **onde Φ é a função de distribuição da N(0,1), que é tabelada. Alguns valores importantes são:**

$\Phi(1.645) = 0.95$, $\Phi(1.96) = 0.975$ e $\Phi(2.326) = 0.99$

Distribuição Normal

- O **Excel** fornece diretamente o valor de $\Phi(z_0)$ através da função **DIST.NORMP**.
- O **único argumento** para esta função é o valor z_0 para o qual você quer calcular a probabilidade de estar abaixo, pois a função pressupõe que a distribuição usada é a Normal padrão (média 0 e variância 1).

Tabela da N(0,1) usando $\Phi(z_0)$

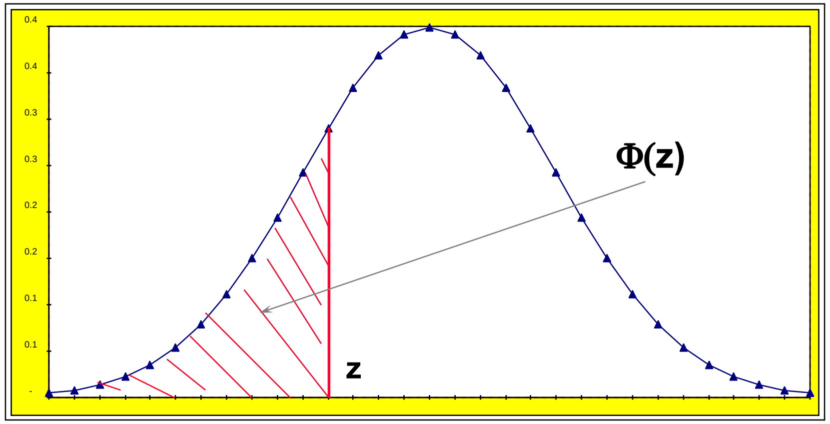


Tabela da N(0,1)

- Simetrias
- $\Phi(-z) = 1 - \Phi(z)$ se $z > 0$
- ISSO É IMPORTANTE POIS A TABELA SÓ CONTÉM VALORES DE z POSITIVOS!
- Probabilidade de um intervalo simétrico em torno de zero
- $\Pr(-t < Z < t) = 1 - 2\{\Phi(-t)\} = 1 - 2\{1 - \Phi(t)\} = 2.\Phi(t) - 1$ onde $Z \sim N(0,1)$

Tabela da N(0,1) ($\Phi(z_0) = \Pr(Z \le z_0)$)

z	ϕ(z)	z	ϕ(z)	z	ϕ(z)	z	ϕ(z)
0.00	0.5000	0.62	0.7324	1.24	0.8925	1.86	0.9686
0.02	0.5080	0.64	0.7389	1.26	0.8962	1.88	0.9699
0.04	0.5160	0.66	0.7454	1.28	0.8997	1.90	0.9713
0.06	0.5239	0.68	0.7517	1.30	0.9032	1.92	0.9726
0.08	0.5319	0.70	0.7580	1.32	0.9066	1.94	0.9738
0.10	0.5398	0.72	0.7642	1.34	0.9099	1.96	0.9750
0.12	0.5478	0.74	0.7704	1.36	0.9131	1.98	0.9761
0.14	0.5557	0.76	0.7764	1.38	0.9162	2.00	0.9772
0.16	0.5636	0.78	0.7823	1.40	0.9192	2.02	0.9783
0.18	0.5714	0.80	0.7881	1.42	0.9222	2.04	0.9793
0.20	0.5793	0.82	0.7939	1.44	0.9251	2.06	0.9803
0.22	0.5871	0.84	0.7995	1.46	0.9279	2.08	0.9812
0.24	0.5948	0.86	0.8051	1.48	0.9306	2.10	0.9821
0.26	0.6026	0.88	0.8106	1.50	0.9332	2.12	0.9830
0.28	0.6103	0.90	0.8159	1.52	0.9357	2.14	0.9838
0.30	0.6179	0.92	0.8212	1.54	0.9382	2.16	0.9846
0.32	0.6255	0.94	0.8264	1.56	0.9406	2.18	0.9854
0.34	0.6331	0.96	0.8315	1.58	0.9429	2.20	0.9861
0.36	0.6406	0.98	0.8365	1.60	0.9452	2.22	0.9868
0.38	0.6480	1.00	0.8413	1.62	0.9474	2.24	0.9875
0.40	0.6554	1.02	0.8461	1.64	0.9495	2.26	0.9881
0.42	0.6628	1.04	0.8508	1.66	0.9515	2.28	0.9887
0.44	0.6700	1.06	0.8554	1.68	0.9535	2.30	0.9893
0.46	0.6772	1.08	0.8599	1.70	0.9554	2.32	0.9898
0.48	0.6844	1.10	0.8643	1.72	0.9573	2.34	0.9904
0.50	0.6915	1.12	0.8686	1.74	0.9591	2.36	0.9909
0.52	0.6985	1.14	0.8729	1.76	0.9608	2.38	0.9913
0.54	0.7054	1.16	0.8770	1.78	0.9625	2.40	0.9918
0.56	0.7123	1.18	0.8810	1.80	0.9641	2.42	0.9922
0.58	0.7190	1.20	0.8849	1.82	0.9656	2.44	0.9927
0.60	0.7257	1.22	0.8888	1.84	0.9671	2.46	0.9931

Tabela da N(0,1)

- Dicas
- Você precisa explorar as simetrias da N(0,1) pois a tabela só é dada para valores positivos de z_0. Por causa da simetria em torno de zero, $\Phi(0) = 0.5$ e $\Phi(z_0)$ é menor que 0.5 se z_0 for um número negativo.

- Se você tiver dúvidas, faça um desenho!
- Lembre-se sempre que $\Phi(z_0)$ é uma função de distribuição, ou seja, mede a probabilidade de estarmos ABAIXO do ponto z_0 .

Distribuição Normal

- Seja $X \sim N(\mu, \sigma^2)$ e $k > 0$. Mostre que $\Pr\{\mu - k\sigma < X < \mu + k\sigma\}$ só depende de k (não depende de μ e σ).
- Solução
- Note que a probabilidade desejada é a probabilidade de X estar a uma distância menor ou igual a k desvios padrões da sua média.

Distribuição Normal

$$\Pr(\mu - k\sigma < X < \mu + k\sigma) = \Pr(-k\sigma < X - \mu < +k\sigma) =$$

$$= \Pr\left(-\frac{k\sigma}{\sigma} < \frac{X-\mu}{\sigma} < +\frac{k\sigma}{\sigma}\right) = \Pr\left(-k < \frac{X-\mu}{\sigma} < k\right) = \Pr(-k < Z < +k) =$$

$$= 2.\Phi(k) - 1$$

- As probabilidades para alguns valores k estão abaixo:

$\Pr(\mu - \sigma < X < \mu + \sigma) = 2.\Phi(1) - 1 = 0.6826$

$\Pr(\mu - 1.645\sigma < X < \mu + 1.645\sigma) = 2.\Phi(1.645) - 1 = 0.90$

$\Pr(\mu - 1.96\sigma < X < \mu + 1.96\sigma) = 2.\Phi(1.96) - 1 = 0.95$

$\Pr(\mu - 2.57\sigma < X < \mu + 2.57\sigma) = 2.\Phi(2.57) - 1 = 0.99$

Distribuição Normal

- Na verdade, aquela "regra de bolso" que diz que 68% dos valores estão a uma distância de 1 d.p. da média e 95% dos valores estão a dois desvios da média acabou de ser mostrada no slide anterior.

- Mas note que isso só é realmente verdade para a distribuição Normal!

Distribuição Normal

- Exemplo

  Numa agência bancária localizada numa grande cidade brasileira, verificou-se que os clientes pessoa física mantêm, em média, um volume de R$ 4800,00 aplicados no banco.

  A dispersão entre os volumes de recursos, medida pelo desvio padrão, é R$ 1600,00. Além disso, pode-se encarar os saldos dos correntistas como independentes entre si e Normalmente distribuídos.

Distribuição Normal

- O banco pretende abrir uma nova agência e seus executivos imaginam que o poder aquisitivo nesta nova área é semelhante ao dos clientes desta agência.

- a) Um cliente é VIP se está entre os 5% com maior volume de recursos. Quanto uma pessoa deveria manter no banco para ser considerada cliente VIP?

Distribuição Normal

- b) O banco pretende cobrar tarifas mais altas dos clientes que têm um baixo volume de recursos aplicados na instituição.

  Os clientes cujos volumes de recursos estão entre os 10% mais baixos terão de pagar esta tarifa mais alta. Abaixo de qual volume um cliente será alvo desta tarifa diferenciada?

Distribuição Normal

- Solução

  Seja X a variável que mede o volume de recursos de um cliente típico da agência. Então X é Normal (4800, $(1600)^2$). Daí: $Z = \frac{X - 4800}{1600}$

  tem densidade Normal padrão.

  Para estar entre os 5% mais "ricos", precisamos encontrar z_0 tal que $\Phi(z_0)$ = 95%. Usando a função INV.NORMP do Excel, encontramos z_0 = 1.645.

  Logo, $$\frac{X - 4800}{1600} = 1.645 \Rightarrow X = 4800 + 1.645(1600) = 7432$$

Distribuição Normal

- Solução (continuação)

  b) Para estar entre os 10% mais "pobres" precisamos encontrar z_0 tal que $\Phi(z_0)$ = 10%. A função INV.NORMP do Excel fornece z_0 = -1.281. Logo,

$$\frac{X-4800}{1600} = -1.281 \Rightarrow X = 4800 - 1.281(1600) = 2750.40$$

- Ou seja, clientes com volume de recursos abaixo de R$ 2750 estarão sujeitos a uma tarifa mais alta, e aqueles com volume de aplicações acima de R$ 7432 terão tratamento VIP.

Distribuição Normal

- Exemplo
- O saldo devedor dos usuários de um certo cartão de crédito é uma variável aleatória Normal com média R$ 200 e desvio padrão R$ 75.
  - a) Qual a probabilidade do saldo devedor de um usuário estar entre R$ 100 e R$ 300?
  - b) Qual deve ser o seu saldo devedor para que você esteja entre os 5% mais endividados?
- Solução

  X é Normal com média 200 e desvio padrão 75 e assim

  Z =(X- 200)/75 é N(0,1).

Distribuição Normal

- Solução (continuação)

  Pr(100 < X < 300) =

$$\Pr\left(\frac{100-200}{75} < Z < \frac{300-200}{75}\right) = \Pr(-1.333 < Z < +1.333) =$$

$$= \Phi(1.333) - \Phi(-1.333) = 2.\Phi(1.333) - 1 = 0.8176$$

  b) Para que você esteja entre os 5% mais endividados, o saldo devedor padronizado deve ser igual a 1.645 (veja tabela da Normal). Daí:

$$Z = \frac{X-200}{75} = 1.645 \Rightarrow X = 200 + 1.645(75) = 323.38$$

  é o saldo para estar entre os 5% com maior saldo devedor.

Distribuição Normal

- Exemplo (para casa)
  - O consumo médio residencial de energia elétrica nos meses de verão numa certa cidade é uma variável Normal com média 210 kWh e desvio padrão 18 kWh.

    a) Qual a probabilidade de que o consumo no verão exceda 225 kWh?

    b) Calcule a probabilidade de que o consumo no verão seja inferior a 190 kWh.

    c) Quanto você deve consumir para estar entre os 2.5% que mais gastam energia?

Distribuição Normal

- **Exemplo (para casa)**
- **Numa certa empresa de informática, o salário *anual* médio dos funcionários com menos de 5 anos de experiência é R\$ 24000, com desvio padrão de R\$ 3000. Suponha que os salários têm distribuição Normal e calcule os valores pedidos a seguir.**

Distribuição Normal

- **a) Qual a probabilidade do salário anual de um funcionário qualquer com menos de 5 anos de experiência ser menor que R\$ 20000?**
- **b) Qual deve ser o valor do salário anual de um funcionário com menos de 5 anos de experiência se 95% dos funcionários (com menos de 5 anos de experiência) tem salário abaixo dele?**

Distribuição Normal

- **c)Toma-se uma amostra de 36 funcionários com menos de 5 anos de experiência. Qual a probabilidade do salário médio na amostra exceder R\$ 24500?**
- **d)Toma-se uma amostra de 12 funcionários com menos de 5 anos de experiência. Qual a probabilidade do maior salário na amostra exceder R\$ 28000?**

Combinações Lineares de Variáveis Normais

- Sejam X_1, X_2, ..., X_n variáveis aleatórias **independentes**, onde $X_i \sim N(\mu_i, \sigma_i^2)$ e seja $Y = X_1 + X_2 + ... + X_n$.
- Então Y tem distribuição Normal com média μ_y e variância σ_y^2 dadas por:

$$\mu_y = \sum_{i=1}^{n} \mu_i$$

$$\sigma_y^2 = \sum_{i=1}^{n} \sigma_i^2$$

Combinações Lineares de Variáveis Normais

- Dois casos particulares importantes são:
  - se os X_i ´s forem iid $N(\mu, \sigma^2)$, então sua soma é Normal com média n. μ e variância n. σ^2 e
  - a média amostral é Normal com média μ e variância σ^2/n.

Distribuição Normal

- Exemplo (continuação)
- Considere o exemplo dos saldos em aplicações bancárias. Suponha que tomamos uma amostra de 16 clientes da agência.
- Qual a probabilidade de que o saldo médio das aplicações dos clientes na amostra exceda R$ 4900?

Seja $\overline{X}$ a média dos saldos dos clientes na amostra.

$\overline{X}$ tem distribuição $N\left(4800, \frac{(1600)^2}{16}\right)$

Distribuição Normal

- Então:

$$\Pr(\overline{X} > 4900) = \Pr\left(\frac{\overline{X} - 4800}{1600/4} > \frac{4900 - 4800}{1600/4}\right) = \Pr\left(Z > \frac{100}{400}\right) =$$

$$= \Pr(Z > 0.25) = 1 - \Phi(0.25) = 1 - 0.599 = 0.401$$

Distribuição Normal (para casa)

- Um estudante universitário gasta em média R$ 600,00 em livros por ano. A dispersão entre os valores gastos, medida pelo desvio padrão, é R$ 240,00.
- Além disso, pode-se encarar os valores gastos pelos universitários como independentes entre si e Normalmente distribuídos. Também, a maioria dos estudantes adquire livros pela Internet.

Distribuição Normal (para casa)

- a) Uma grande livraria na Internet pretende oferecer um cartão VIP aos clientes que mais compram livros. Apenas os 1% que mais consomem livros num período de um ano receberão o cartão. Acima de qual volume anual de compras um consumidor se candidata ao cartão VIP?
- b) Considere 16 estudantes universitários. Qual a probabilidade do gasto médio anual em livros destas 16 pessoas ultrapassar R$ 660,00?
- c) Dentre as 16 pessoas nesta mesma amostra, qual a probabilidade do estudante que menos consumiu livros ter gasto mais de R$ 650 no ano?

Distribuição Normal (para casa)

- Um apartamento de 2 quartos numa certa região da cidade custa, em média R$ 260 mil. A dispersão entre os valores, medida pelo desvio padrão, é R$ 100 mil.
- Além disso, pode-se encarar os preços dos apartamentos como independentes entre si e Normalmente distribuídos.

Distribuição Normal (para casa)

- a) Uma imobiliária pretende oferecer uma viagem de presente aos compradores de apartamentos de 2 quartos neste bairro que comprem os apartamentos situados na faixa dos 10% mais caros. A partir de quanto deve custar o seu apartamento para que você ganhe a viagem de "presente"?
- b) Considere 16 compradores de apartamentos de 2 quartos neste bairro. Qual a probabilidade do preço médio pago por eles ser inferior a R$ 300 mil?
- c) Dentre as 16 pessoas nesta mesma amostra, qual a probabilidade do comprador que pagou mais caro por um apartamento ter pago menos de R$ 285 mil?

A distribuição Lognormal

- A distribuição Lognormal é uma distribuição de probabilidade contínua usada para dados positivos.
- Esta distribuição é freqüentemente usada na modelagem do preço de ações e outros ativos financeiros, e também pode modelar o tempo até a ocorrência de um defeito de uma máquina.

A distribuição Lognormal

- Veja o link: http://www.inf.ethz.ch/personal/gut/lognormal/ para um simulador interessante de variáveis lognormais e normais.
- Se você se interessar, o artigo do link:
- http://stat.ethz.ch/~stahel/lognormal/bioscience.pdf
- discute o uso da lognormal nas ciências.

A Distribuição Lognormal

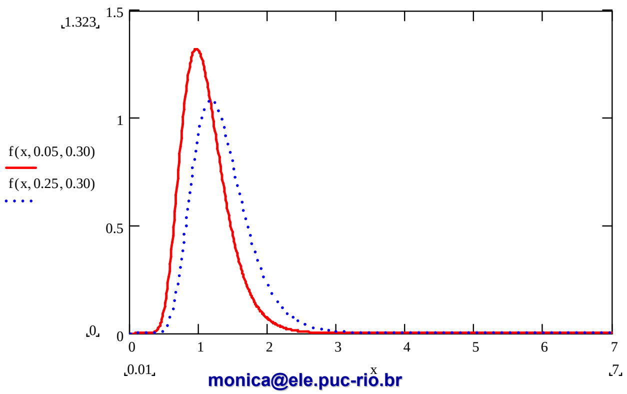
- Como criar uma variável lognormal?
- Seja $X \sim N(\mu, \sigma^2)$. Seja $Y = \exp(X)$. Então Y tem densidade Lognormal com parâmetros μ e σ^2.

- A densidade de Y é dada por:

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \left(\frac{1}{y}\right) \cdot \exp\left(-\frac{(\log(y)-\mu)^2}{2\sigma^2}\right) \quad \text{onde } \; y > 0$$

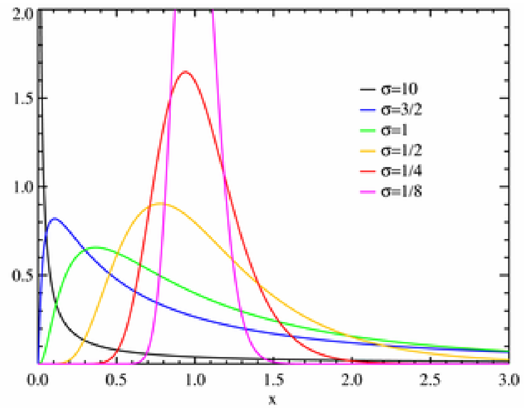
A Distribuição Lognormal

- Densidades Lognormais com μ = 0.05 e 0.25 e σ = 0.30



A Distribuição Lognormal

- Densidades Lognormais com μ = 0 e diversos valores para σ.



A Distribuição Lognormal

- *Atenção:*
- *A distribuição Lognormal, ao contrário do que o nome indica, não significa a densidade do logaritmo de uma variável Normal, pois uma variável Normal admite valores negativos, onde o logaritmo não está definido. Uma variável aleatória com densidade Lognormal é encontrada tomando-se a exponencial de uma variável aleatória Normal!*

A Distribuição Lognormal

- **A densidade Lognormal pode ser pensada como gerada pelo PRODUTO de diversos fatores que são todos independentes entre si.**
- **Por que? Pois Y = exp(X) e X é Normal, que pode ser encarada como a soma de fatores independentes (é a idéia do Teorema Central do Limite). Ao exponenciarmos, esta soma torna-se um produto...**

Lognormal como modelo para o preço de uma ação

- **Uma forma de descrever a incerteza sobre o preço de uma ação é supor que as variações no preço entre os instantes t e t+Δt podem ser divididas em 2 componentes, uma aleatória e a outra determinística, como a seguir:**

$$S_{t+\Delta t} = S_t.\left\{\exp\left(\mu.\Delta t + \sigma.Z\sqrt{\Delta t}\right)\right\}$$

- **onde Z é uma variável N(0,1) e μ e $\sigma > 0$ são parâmetros conhecidos. O parâmetro μ representa a taxa média de crescimento do preço ao longo do tempo.**

Lognormal como modelo para o preço de uma ação

- **Note que, se $\sigma = 0$, a evolução dos preços é puramente determinística, e então:**

$$S_{t+\Delta t} = S_t.\left\{\exp\left(\mu.\Delta t\right)\right\}$$

- **Nesta expressão percebemos que a tendência determinística dos preços é crescente desde que $\mu > 0$.**
- **Se $\sigma > 0$ então existe uma componente aleatória no comportamento dos preços. Esta componente aleatória é dada por uma variável aleatória N(0,1), e assim o efeito desta variável pode ser o de atenuar o crescimento determinístico no preço, pois Z pode ser negativo. Note que a variável exp(Z) é Lognormal.**

Média e variância da Lognormal

- Se Y ~ Lognormal(μ, σ^2) então:

  E(Y) = exp(μ + σ^2/2)

$$VAR(Y) = \exp\left(2\mu + \sigma^2\right).\left(e^{\sigma^2} - 1\right)$$