Avaliação Econômica de Ativos Permanentes

Prof. Dr. Marco Antonio Leonel Caetano

TÓPICO 1

A Arte da Compreensão de Incertezas utilizando Estatística Descritiva

1.1 Introdução

Na divulgação de toda avaliação econômica, pesquisadores se deparam com o fato de como tratar e qual a melhor forma de apresentação de dados obtidos através dos experimentos.

Necessariamente toda divulgação deve sempre começar pela divulgação dos resultados utilizando técnicas estatísticas adequadas. Dentro da Estatística, existe um amplo campo de técnicas , para melhor expressar e representar resultados de experimentos dentro de três sub-áreas bastante distintas, porém muito bem conectadas: Estatística Descritiva, Probabilidade e Estatística Indutiva. No campo da Estatística Descritiva serve como objeto de utilização, ferramentas como gráficos, tabelas e análises de formas e estruturas das representações dos dados. Este tipo de Estatística não tem valor de inferência, ou seja, não se pode concluir ou expressar conclusões de um experimento só com este tipo de técnica.

No campo da Probabilidade, teoremas garantem os tipos de distribuições que regem um experimento, quais são as mais adequadas e com que confiança os dados poderão ser coletados de forma a serem representativos de uma população. Finalmente, no último campo, o da Estatística Indutiva ou mais conhecido como Inferência, técnicas garantem as conclusões com grande acurácia indicando os erros e confiabilidades dos resultados obtidos. Também nesse campo, pode-se fazer previsão de tendências e correlações entre as variáveis e parâmetros obtidos experimentalmente.

Assim, realizar um experimento sem o devido cuidado com seu tratamento e forma de apresentação pode comprometer todo trabalho por falta de compreensão ou de interpretações errôneas sobre determinados resultados financeiros.

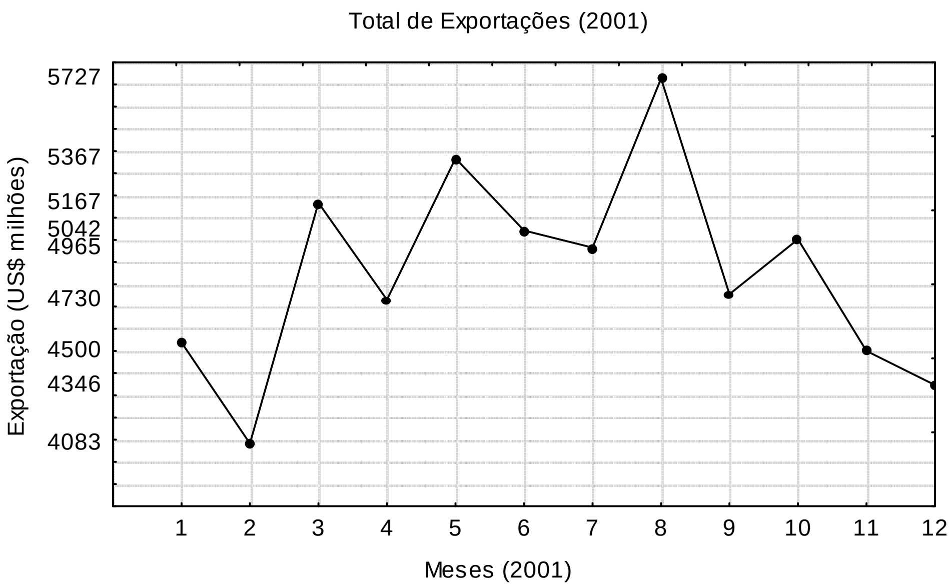


Figura 1.1: Gráfico de Linhas - Exportação Brasileira (2001)

1.2 Representação Gráfica

A melhor forma de representação de dados é sob gráficos, que muitas vezes por si só são bastante explicativos e até conclusivos dependendo do tipo de trabalho. Para expressar alguns tipos de gráficos, como exemplo, os dados abaixo representam as exportações brasileiras em US$ milhões no ano de 2001, mês a mês:

Tabela 1.1 - Exportações (total em US$milhões)

Mês	1	2	3	4	5	6	7	8	9	10	11	12
US$	4538	4083	5167	4730	5367	5042	4965	5727	4755	5003	4500	4346

Fonte: Ministério do Desenvolvimento, Indústria e Comércio Exterior

1.2.1 Gráfico de Linhas

Esta é uma das formas mais simples de apresentação. O gráfico é apresentado com a união simples de retas entre os pontos do experimento. É muito útil principalmente quando se quer, em primeira instância, verificar tendências dos resultados obtidos. A figura 1.1 apresenta um gráfico típico de linhas.

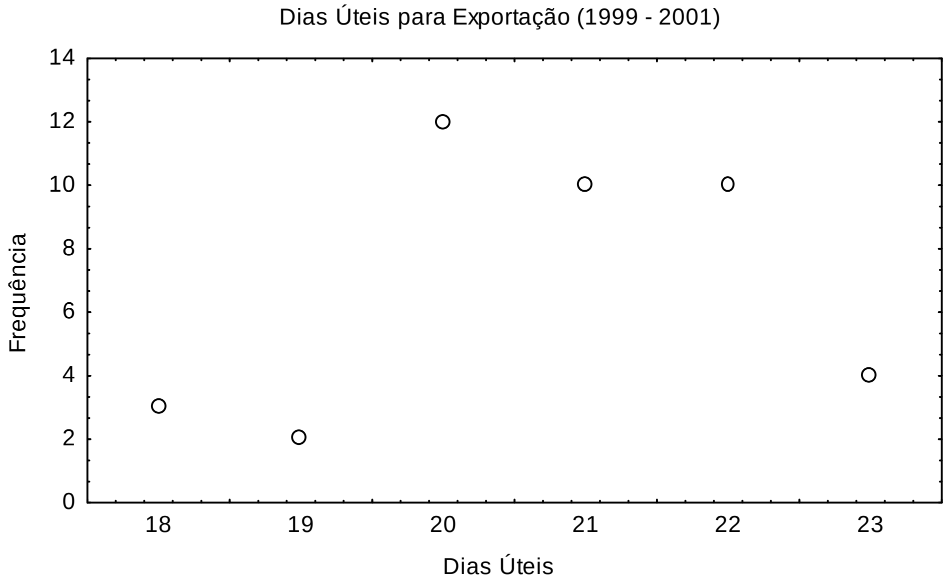


Figura 1.2: Gráfico de Pontos - Dias Úteis

1.2.2 Gráfico de Pontos

Esta forma de gráfico é bastante interessante quando se observam freqüências nos dados coletados. Neste caso, o leitor consegue visualizar o valor que mais se repete em uma amostragem. Este tipo de gráfico serve também para representar séries econômicas históricas.

Os dados a seguir são referentes à freqüência observada nos dias úteis mensais para exportação brasileira entre 1999 e 2001.

Tabela 1.2 - Freqüência de Dias Úteis Mensais (1999 - 2001)

Dias Úteis	Freqüência
18	3
19	2
20	12
21	10
22	10
23	4

Fonte:Ministério do Desenvolvimento, Indústria e Comércio Exterior

O gráfico então é feito colocando-se pontos nos valores observados, seguindo no eixo x e na freqüência com que aparecem na amostra no eixo y. A figura 1.2 mostra o gráfico de pontos.

1.2.3 Histograma

O histograma consiste em retângulos justapostos indicando em sua base o intervalo dos valores de dados do experimento cuja freqüência é representada pela altura do retângulo. O sentido é um pouco mais amplo do que o gráfico de pontos, pois o interesse neste caso não é quanto a um único valor, mas com relação a um intervalo de valores amostrados.

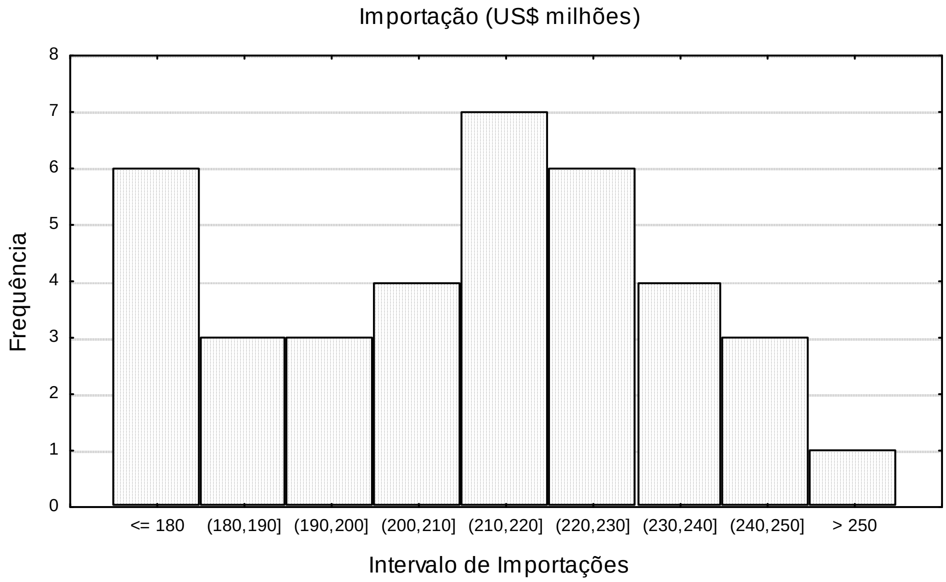


Figura 1.3: Histograma dos Valores de Importação Brasileira (1999-2002)

Os valores a seguir são correspondentes à média diária mensal de importação brasileira de Janeiro de 1999 a Janeiro de 2002.

Tabela 1.3 - Média Diária Mensal de Importação (US$ milhões)

183	175	176	183	194	212	183	203	202	223
226	193	170	192	212	210	213	219	232	235
252	247	252	243	228	222	247	230	234	238
220	221	218	216	210	174	172			

Fonte: Ministério do Desenvolvimento, Indústria e Comércio Exterior

O histograma representando os valores da Tabela 1.3 é apresentado na Figura 1.3 e pode-se notar alguns fatos interessantes. Por exemplo, ele auxilia a interpretar que o valor mais freqüênte de Importação está entre US$210 e US$220 milhões de média diária.

1.2.4 Gráfico de Barras

Assim como o histograma representa os valores obtidos no experimento em em termos de frequência para cada valor observado. A diferença é que não se utiliza este gráfico para intervalos amostrados, mas para os valores observados de maneira individual. A Figura 1.4 apresenta a representação dos valores para a importação brasileira da Tabela.1.3 no Exemplo 3.

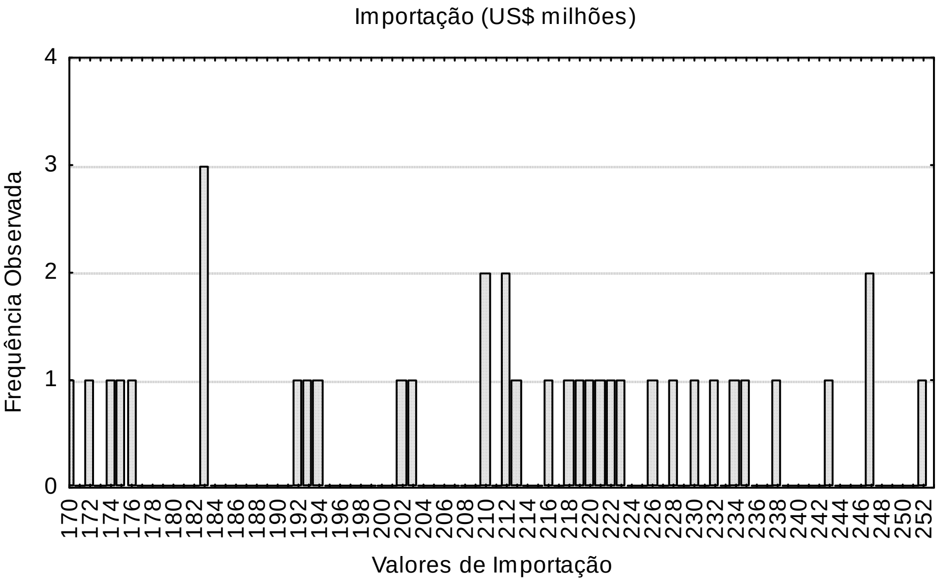


Figura 1.4: Gráfico de Barras - Importação Brasileira

1.2.5 Curvas de Nível

Este é um tipo bastante interessante de gráfico pois traça isolinhas para os pontos amostrados. Isto significa que uma vez o valor escolhido numa das linhas, percorrendo essa linha, todos os pontos para as posições x e y são iguais. A curva representa uma função em 3 dimensões como se fosse uma foto bidimensional de um terreno em 2D.

Exemplo 1 *A Tabela 1.4 a seguir apresenta os valores da execução orçamentária das despezas federais de Janeiro a Setembro de 1996 - 1999, para a Administração Federal e Saúde. Supõe-se que uma curva de ajuste boa para a relação gastos com administração x gastos com saúde seja,*

$$z = x^2 + y^2$$

onde x aqui é administração e y é saúde. Os dados reais são apresentados com os valores do orçamento em milhões de reais. Os dados foram obtidos do site do IPEA composto de medidas mensais da fonte do governo federal SIAF-CCONT / STN. As isolinhas para esses valores são as formas traçadas no gráfico da Figura 1.5.

Tabela 1.4 - Orçamento Federal (R$ milhões)

Administração	Saúde
6.532	13.219
18.091	11.836
19.442	10.986
19.740	10.220

Fonte: SIAF - CCONT / STN

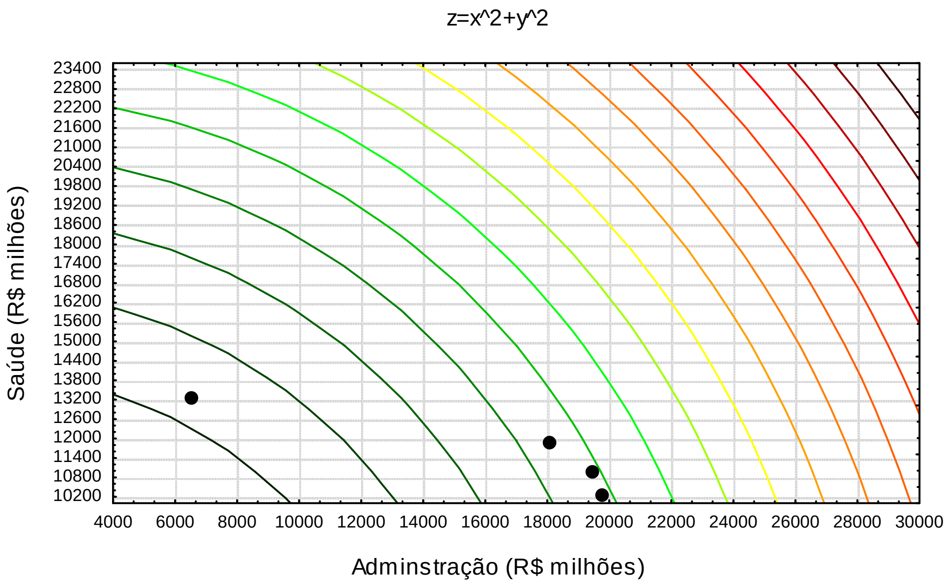


Figura 1.5: Isolinhas dos gastos federais

1.3 Medidas Descritivas dos Dados

Na seção anterior, foram apesentadas formas gráficas de representação dos dados de um relatório empresarial. Cabe ao gestor escolher e adequar a melhor forma de apresentação de seus resultados de forma a elucidar todos os fatos com uma simples visualização dos acontecimentos. No entanto, na maioria das vezes essa facilidade não é obtida e por várias razões. Seja pela complexidade do fenômeno, seja pela modelagem com um número extremamente grande de variáveis ou parâmetros, a simples escolha de um tipo de gráfico não consegue expressar quantitativamente a importância de certas relações existentes. Neste ponto cabe então fazer uso de formas quantitativas de extração de informações, através de medidas estatísticas que apresentem de forma rápida e suscinta as inter-relações existentes no fenômeno em estudo. Então, o gestor deve fazer uso de variável, como forma de representação genérica dos principais fatores decorrentes do experimento.

Uma variável pode ser discreta ou contínua, dependendo do tipo de estudo executado. Variável discreta é toda aquela relacionada a números inteiros, ou seja, entre um período ”t” de observação e outro ”t+1”, não se encontra valores amostrados. Normalmente esse tipo de variável é utilizado em problemas de contagem. Exemplos disso são contagens de firmas em concordata, nível de emprego, contagem do número de vagas abertas por uma empresa, etc.

Para uma variável contínua, como o próprio nome diz, os dados podem até serem observados de forma discreta, mas as relações empresariais por exemplo acontecem continuamente. Essas variáveis são sempre representadas por números reais. Um exemplo de variáveis contínuas é apresentado na Tabela 1.5.

Exemplo 2 *Apesar da taxa de desemprego no Brasil ser uma medida semanal ou mensal (em %) pode ser considerada como uma medida contínua no tempo, pois seus valores são números reais.*

Tabela 1.5 - Taxa de Desemprego No Brasil (Jan/1999 a Maio/2001, %)

7,73	7,51	8,16	8,02	7,70
7,84	7,54	7,68	7,37	7,53
7,32	7,30	7,60	8,20	8,10
7,80	7,80	7,40	7,20	7,10
6,70	6,80	6,19	4,83	5,70
5,73	6,46	6,51	6,86	

Fonte: SEADE

O primeiro tratamento representativo para extração de informação dessa coleta é através de uma tabela, conhecida como tabela de classes. Nesse tipo de tabela deseja-se informar a variação dos dados separados em classes de importância e não de maneira isolada. Assim, algumas definições precisam ser colocadas.

(i) **Dados Brutos (n)** - Dados ainda não organizados, como na Tabela 1.5.
(ii) **Rol** - É o arranjo dos dados brutos em ordem crescente ou decrescente.
(iii) **Range ou Amplitude Total** - É a diferença entre o maior e o menor valor observado.
(iv) **Freqüência Absoluta da Classe (F)** - Número de vezes que o elemento aparece na amostra ou o número de elementos pertencentes a uma classe.
(v) **Número de Classes (k)** - existem duas maneiras para determinar um número adequado de classes.
(a) Número será k = 5 se o número de dados for menor ou igual a 25.
(b) Para número de dados superior utiliza-se

$$k = \sqrt{n}$$

(vi) **Amplitude das Classes (h)** -

$$h = \frac{Range}{k}$$

(vii) **Limite das Classes** - Li : Limite Inferior ; Ls : Limite Superior

Li j———j Ls : Comprende os valores Li e Ls
Li j——– Ls : Não compreende o valor Ls
Li ———j Ls : Não Compreende o valor Li

(viii) **Pontos Médios das Classes (PMi) -** É a média dos valores limtantes das classes.

$$PM_i = \frac{L_i + L_s}{2}$$

(ix) **Freqüência Absoluta Acumulada Direta (Fac) -** É a soma das freqüências absolutas dos valores inferiores ou igual ao valor da freqüência da classe.
(x) **Freqúência Relativa (f) -** Porcentagem do número de dados da classe em relação ao total de dados.

$$f = \frac{F}{n}$$

Uma vez colocadas essas definições, os 29 dados brutos da Tabela 1.5 podem informar melhor segundo a Tabela 1.6 (tabela de classes) para o nível de desemprego no país.

Tabela 1.6 - Tabela de Classes para Nível de Desemprego no Brasil (Jan/1999 a Maio/2001)

Classes	F	f	PM	Fac
4,83|------5,504	1	0,034 (3,4%)	5,16	1
5,504|-----6,178	2	0,068 (6,8%)	5,60	3
6,178|-----6,852	5	0,172 (17,2%)	6,51	8
6,852|-----7,526	8	0,275 (27,5%)	7,18	16
7,526|-----8,20	13	0,448 (44,8%)	7,86	29
total	29	1 (100%)		29

Fonte: SEADE

Essa tabela é bastante útil na construção do histograma e mostra qual a classe de concentrações (em percentagem) mais freqüentes de desemprego. Pode-se observar que a maior freqüência de percentagem ocorre para as classes entre 7,526% a 8,20 %, de desemprego o que corresponde a 44,8% dos dados (freqüência relativa).

1.3.1 Medidas de Posição

As medidas de posição são definidas de modo a apresentar o valor em torno do qual os dados se distribuem. Essas medidas são também conhecidas como medidas de tendência central pois estabelecem uma indicação do elemento central da amostragem realizada. As principais medidas são a média, mediana e moda.

Média Aritmética

(i) **Dados não agrupados**. Sejam x_1; x_2; ...; x_n valores da variável x. A média aritmética para os dados brutos, coletados em um experimento será:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

(ii) **Dados agrupados em tabela de freqüência**. Sejam x_1; x_2; ...; x_n com freqüências F_1; F_2; ...; F_n respectivamente. Assim a média será:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i F_i}{n}$$

Exemplo 3 *A Tabela 1.7 representa o número de cheques sem fundos devolvidos na segunda vez em cada 1000 cheques apresentados, de Maio de 2001 a Maio de 2002.*

Tabela 1.7 - Cheques Sem Fundo (média /1000)

Devolução	Freqüência Absoluta
14,1	4
13,7	2
13,6	2
14,5	2
16,2	1
14,9	1
12,6	1

Fonte: Serasa

A média aritmética no caso para a Tabela-1.7 é 14,1 cheques entre Maio de 2001 e Maio de 2002 para cada 1000 apresentados.

(iii) **Dados Agrupados em Tabela de Classes**. As Classes são representadas pelos seus pontos médios, conforme a Tabela 1.6. Neste caso a média é calculada

$$\bar{x} = \frac{\sum_{i=1}^{n} PM_i F_i}{n}$$

Observando a Tabela de Classes 1.6, pode-se calcular a sua média pela fórmula anterior, a qual fornece o valor médio $\bar{x} =$ 7,19.

Mediana

Um valor é dito mediano, quando ele divide o conjunto de dados do experimento em dois subconjuntos com igual número de elementos. Sua notação em geral é $\tilde{x}$:

(i) **Dados não agrupados**. Sejam os dados

5 7 8 10 14 A mediana é $\tilde{x}$ = 8

5 7 8 10 14 15 A mediana é $\tilde{x}$ = 9

Assim, uma maneira de se encontrar a mediana de um conjunto composto por dados brutos seria da seguinte forma. Se o número de dados "n" é ímpar, a mediana é o elemento central (n+1)/2, caso contrário a mediana será a média dos elementos centrais formados por [n/2 , (n/2)+1].

(ii) **Dados agrupados por freqüência**. Neste caso, cria-se uma nova coluna das freqüências acumuladas diretas para auxílio na escolha da mediana. A Tabela 1.7 passaria para a forma da Tabela 1.8 a seguir.

Tabela 1.8 - Cheques Sem Fundo (média /1000)

Devolução	Freqüência Absoluta	Freqüência Relativa
14,1	4	4
13,7	2	6
13,6	2	8
14,5	2	10
16,2	1	11
14,9	1	12
12,6	1	13

Fonte: Serasa

Neste caso n=13 é ímpar. Tem-se então neste caso que (n + 1)/2=7, o que significa que o sétimo elemento corresponde ao elemento mediano desse conjunto de valores de cheques devolvidos. Logo a mediana será 13,6 diferente da média que é de 14,1.

(iii) **Dados Agrupados em Tabela de Classes**. Neste caso será necessário uma fórmula de interpolação para se encontrar o elemento mediano. Deve-se ressaltar que esse valor é apenas representativo e que não necessariamente fará parte da amostra. Os passos a seguir serão:
(a) Calcula-se a ordem (n/2) não se preocupando se for par ou ímpar pois a variável será contínua. A classe da mediana é aquela cuja freqüência acumulada até ela é maior ou igual a n=2 e a imediatamente anterior menor que n/2.
(b) Utiliza-se a seguinte fórmula de interpolação:

$$\tilde{x} = L_{md} + \frac{\left(\frac{n}{2} - \sum f\right)}{F_{md}} \times h$$

onde
L_{md} : limite inferior da classe da mediana.
n : tamanho da amostra
$\sum f$: freqüência acumulada da classe imediatamente anterior à da mediana.
h : Amplitude da classe da medina
F_{md} : freqüência absoluta da classe da mediana.

Assim, como exemplo, observando a Tabela 1.6 a classe da mediana seria a quarta classe, ou seja,

6,852 |------------- 7,526

uma vez que o décimo quarto elemento (n/2) pertence a essa classe. Então o cálculo da mediana seria:

$$\tilde{x} = 8 + \frac{(14{,}5 - 8)}{16} \times (0{,}674) = 8{,}27$$

Existem ainda medidas alternativas para se dividir os dados em quatro partes iguais, dez e cem partes. Elas são conhecidas como Quartis, Decis e Percentis respectivamente. A única alteração na fórmula é a troca de n/2 por n/4 no caso de Quartis, n/10 no caso de Decis e n/100 no caso de percentis. Os limites e as freqüências acumuladas diretas também são trocados pelos limites das classes dos Quartis, Decis e Percentis.

Moda

Essa medida representa o elemento mais freqüênte na amostragem, ou seja, aquele que mais se repete.

(i) **Dados não agrupados.**

Exemplo 4 *Sejam os dados de uma amostragem composta por 2,7,9,5,6,3,7,4,1,7. A moda neste caso é o número 7.*

(ii) **Dados em Tabela de Classes.** Da mesma forma que na mediana faz-se necessária a interpolação dos dados para encontrar a moda. Deve-se seguir os seguintes passos:
(a) identifica-se a classe modal, ou seja, aquela que tenha a maior freqüência absoluta.
(b) Utiliza-se a fórmula:

$$Mo = L_i + \frac{(F_i - F_{i-1})}{2F_i - F_{i-1} - F_{i+1}} \times h$$

Li : limite inferior da classe modal.
Fi : freqüência absoluta da classe modal
F_{i-1}: freqüência absoluta da classe imediatamente anterior à classe modal.
F_{i+1}: freqüência absoluta da classe imediatamente posterior à classe modal.
h : amplitude da classe modal.

Novamente observando a Tabela 1.6 de classe pode-se observar que a classe modal é a última classe com 13 elementos. Então aplicando-se a fórmula da moda tem-se:

$$Mo = 7{,}526 + \frac{(13 - 8)}{2 \times 13 - 8 - 0} \times 0{,}674 = 7{,}713$$

1.3.2 Medidas de Dispersão

Uma vez conhecida as medidas de posição de uma curva representativa dos dados de uma avaliação financeira ou empresarial, faz-se necessário saber se esta coleta é representativa da população de dados em estudo ou não. Torna-se indispensável então, o conhecimento da dispersão desses dados em relação as medidas de posição, principalmente em relação a média. São quatro as medidas a serem apresentadas.

Amplitude Total

Essa medida é muito simples e constitui na primeira avaliação sobre a natureza da amostragem. A amplitude total é a diferença entre o maior valor e o menor valor dos dados coletados. Sua utilização é bastante limitada pois apenas depende da dispersão dos valores extremos, não sendo afetada pela dispersão dos valores internos.

Variância

A variância mede a dispersão dos dados em torno da média. A título de exemplo, suponha-se que se tem o seguinte conjunto de dados

A = {3, 4,5,6,7}

onde a média desse conjunto é 5. Calculando-se o desvio das unidades do conunto A em relação à média tem-se:

$$d_1 = x_1 - \bar{x} = -2$$
$$d_2 = x_2 - \bar{x} = -1$$
$$d_3 = x_3 - \bar{x} = 0$$
$$d_4 = x_4 - \bar{x} = 1$$
$$d_5 = x_5 - \bar{x} = 2$$

Esta soma de desvios poderia servir como medida de dispersão não fosse o seguinte fato em que $\sum_{i=1}^{5} d_i = 0$. Ou seja, todas as diferenças dos dados de uma amostra em relação ao elemento central se anulam. Elevando-se esses desvios ao quadrado para eliminar este problema e somando-os tem-se:

$$sqd = \sum_{i=1}^{5} d_i^2 = \sum_{i=1}^{5} (x_i - \bar{x})^2$$

Acontece que como está, essa medida cresceria indefinidamente a medida que novos dados fossem sendo coletados. Logo, para que esse valor não se torne indefinidamente crescente, pondera-se a medida *sqd*, dividindo-a pelo número de dados, ou seja,

$$sqd = \sigma^2 = \sum_{i=1}^{5} \frac{(x_i - \bar{x})^2}{5}$$

Esta forma de medida passa a ser chamada então de variância populacional, uma vez que foi ponderada por todos os termos amostrados. As vezes, nossa intuição em coletar dados nos trai em favor de alguns pontos mais favoráveis conhecidos como viés ou tendenciosidade na amostragem. Uma primeira medida de correção a se fazer é dividir as somas dos desvios não pelo total ”n” de dados, mas por ”n-1” dados. A teoria de probabilidade nos prova que este é um bom ”truque” de correções de tendenciosidade na amostragem. Logo, a segunda medida de variância será:

$$s^2 = \sum_{i=1}^{5} \frac{(x_i - \bar{x})^2}{5-1}$$

onde a nova medida passa a ser chamada de variância amostral. De modo geral pode-se então afirmar que para um conjunto de n dados, tem-se os dois tipos de variâncias:

- **variância populacional**

$$\sigma^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n}$$

- **variância amostral**

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n-1}$$

A variância amostral para o conjunto A descrito anteriormente será 2,5. No caso de se ter dados já apresentados em tabela de freqüências, o cálculo da variância pode ser realizado diretamente através de:

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2 \times F_i}{n-1}$$

onde a variável Fi representa a freqüência absoluta dos dados.

Exemplo 5 *A tabela 1.7 apresenta a devolução de cheques em tabela de freqüência. Para este exemplo a média encontrada foi 14,1 e neste caso a variância amostral pode ser calculada da seguinte forma:*

$$\begin{aligned} s^2 &= \frac{1}{13-1}\Big[(14{,}1-14{,}1)^2 \times 4 + (13{,}7-14{,}1)^2 \times 2 + (13{,}6-14{,}1)^2 \times 2 + (14{,}5-14{,}1)^2 \times 2 + \\ &\qquad + (16{,}2-14{,}1)^2 \times 1 + (14{,}9-14{,}1)^2 \times 1 + (12{,}6-14{,}1)^2 \times 1\Big] \\ &= \frac{8{,}44}{12} \\ &= 0{,}703 \end{aligned}$$

A última forma da apresentação da variância é quando se tem os dados em forma de tabela de classes. Neste caso o cáclulo da variância será:

$$s^2 = \sum_{i=1}^{n} \frac{(PM_i - \overline{x})^2 \times F_i}{n-1}$$

onde PMi é o ponto médio de cada classe.

Exemplo 6 *Utilizando-se da Tabela 1.6, foi encontrado na seção anterior a média para a tabela de classe de 7,19% de nível de desemprego. A variância amostral para este exemplo é 0,618.*

Desvio-Padrão

Esta medida fornece ao pesquisador uma maneira de saber matematicamente a oscilação em torno dos dados. O desvio-padrão fornece qual o grau de confiabilidade dos dados em torno da média. Sabe-se da teoria da Probabilidade que se um conjunto de dados contínuos para ser considerado como um conjunto de dados com distribuição Normal, 68% dos dados devem estar em torno da média no intervalo [Média-Desvio-padrão; Média+Desvio-padrão]. Assim, o desvio-padrão é a raiz quadrada da medida da variância, ou

$$dp = \pm\sqrt{s^2}$$

Coeficiente de Variação

Essa é uma medida relativa da dispersão, ou seja, em porcentagem quanto a variabilidade influencia na confiança da média calculdada. Com um coeficiente de alto grau (por exemplo acima de 50%) não se pode dizer que a média encontrada é representativa para a amostragem realizada. Assim, uma maneira de calcular o coeficiente de variação é relativizar o desvio padrão em relação à média:

$$cv = \frac{s}{\bar{x}}$$

Exemplo 7 *Para a tabela 2.6, pode-se saber se a média encontrada de nível de desemprego é representativa. A média foi de 7,19% de desemprego. Sendo o desvio padrão ±0,786 o coeficiente de variação será:*

$$cv = \frac{0{,}786}{7{,}19} = 0{,}109$$

ou 10,9% de variação. Assim, agora pode-se concluir que a média encontrada é representativa para os dados a respeito do nível de desemprego no país.O *cv* indica que existe uma variabilidade de cerca de 11% ao valor encontrado pela média.

1.3.3 Medidas de Assimetria

Este tipo de medida é bastante útil quando se deseja saber a forma da curva que os dados da amostra se assemelham. Esta curva pode ser simétrica quando a área em relação as medidas de posição são iguais, tanto a frente quanto atrás da média, mediana e moda. Quando essas áreas são diferentes, diz-se que a curva é assimétrica. Essa assimetria será positiva quando o coeficiente de assimetria (AS) é postivo, indicando que o valor modal é inferior ao valor médio. A assimetria será negativa quando o valor modal for maior que o valor médio. O coeficiente de assimetria pode ser calculado como

$$AS = \frac{\bar{x} - Mo}{s}$$

onde o x é o valor médio, Mo o valor modal e s o desvio-padrão. No entanto essa fórmula as vezes pode apresentar um inconveniente. Muitas vezes não se tem um valor modal, ou se tem muitos valores modais. Nestes casos, uma fórmula alternativa é o coeficiente de Pearson que faz uso do valor mediano e dos Quartis na forma,

$$AS = \frac{Q_3 + Q_1 - 2\tilde{x}}{Q_3 - Q_1}$$

onde os Q representam os Quartis terceiro e primeiro e o valor mediano é representado pelo símbolo $\tilde{x}$.

TÓPICO 2
Avaliação Econômica através da Simulação Computacional de Incertezas

2.1 Introdução

O Cálculo de Probabilidades é um ferramental matemático que se presta ao estudo de fenômenos aleatórios ou probabilísticos. A estatística em si apresenta não somente metodologias de medidas de informação, como auxilia na projeção de resultados de eventos com risco mínimo em tomadas de decisão. No entanto, para que esse risco na realidade corresponda a teoria estatística, é necessário que todas as premissas básicas sejam atendidas. Entre essas premissas se destaca como principal, o tipo de coleta de dados e a quantidade de dados amostrados. Em linguagem mais atual, tudo sobre previsão só se prestará como ferramenta importante, se uma empresa possuir um banco de dados perfeitamente correlacionado aos fatos e fluxos de informações das empresas.

Um importante fator na tomada de decisão empresarial se refere ao fato sobre o preço a ser dado para um determinado ativo, após decorrentes algum período de tempo. Em outras palavras, o que se busca é uma quantificação sobre a depreciação existente de um ativo.

Nesse aspecto, com o grande volume de operações financeiras realizadas nos dias atuais, graças aos computadores e operações em *web* tais como *e-business* e *e-commerce*, a estatística tradicional perde totalmente sua importância, se desvinculada de técnicas e operações computacionais.

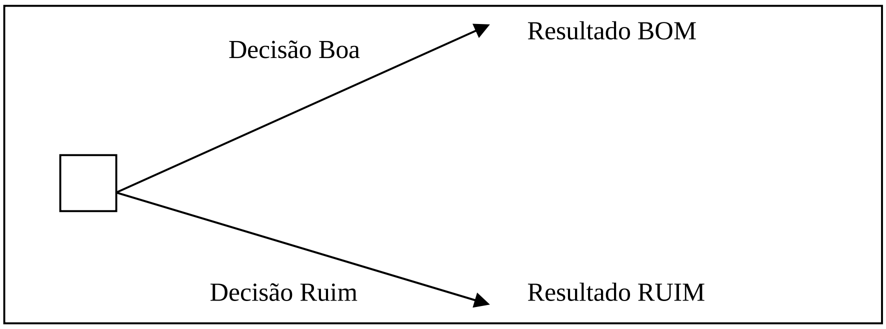
O quanto se paga por uma informação em tempo real (*on line*)? Por que se paga por uma informação em tempo real? Como será possível verificar nesse tópico, quanto menores os tempos de aquisição de uma informação, bem como, quanto maiores os volumes dessas informações, maiores serão as chances de realização de negócios lucrativos com ótimas tomadas de decisões.

2.2 Tomada de Decisão

Todas as pessoas tomam decisões diariamente e de diversas maneiras. Algumas decisões são calculadas, outras são mais irracionais, outras aleatórias. O fato de tomar decisão não significa que de alguma maneira, tomou-se a decisão correta. Decisões sempre tem conseqüências e elas podem ser muito importantes. Problemas de decisões tornam-se cada vez mais difícil se eles são complexos, no sentido de envolver muitas variáveis, ou se requererem múltiplas decisões sucessivas sendo que cada decisão possivelmente tem mais de um resultado. Um problema de tomada de decisão pode ser composto de decisão com certeza e decisão com incertezas nas variáveis que o definem. Numa decisão com certeza, quando se escolhe a alternativa a ser tomada, já se sabe de antemão se o resultado final é bom ou não. Um exemplo de decisão com certeza é quando recebe um prêmio. Por exemplo, se alguém lhe oferece como um prêmio um automóvel e você se recusa, certamente você saberá que essa decisão não foi tão boa. No entanto se recusa em troca do valor do carro em dinheiro (supondo que lhe pagarão o valor de mercado do automóvel) certamente será uma boa decisão (desde que não deseje gastar dinheiro com gasolina, mecânico, etc).

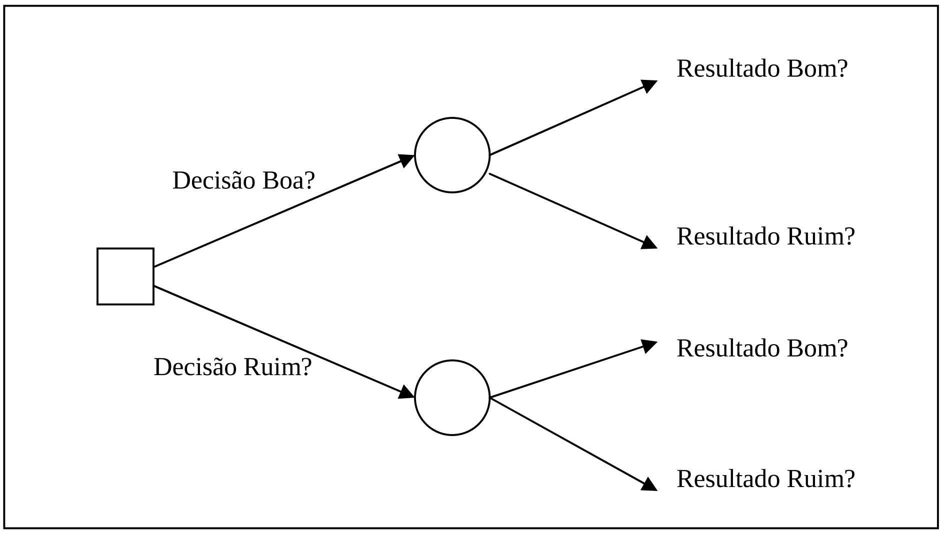
Mas na vida, no dia a dia, dificilmente nos encontraremos a tomar decisões com certeza. O mais comum é sempre depois de uma decisão nos perguntar se ela foi boa ou não.

DECISÃO COM CERTEZA



Esse tipo de tomada de decisão onde a pergunta final é se ela foi acertada ou não, consiste na tomada de decisão com incertezas. Na presença de incertezas sobre o resultado final, o decisor é de fato, forçado a se portar diante de um jogo. No caso de investidores, ou gestores de grandes firmas, sempre estarão diante desse cenário, seja para compra de ações, aplicações de recursos orçamentários, aquisições de imóveis ou mesmo parcerias para aumentar o poder de domínio do mercado.

DECISÃO COM INCERTEZA



Os gráficos anteriores são representado por quadrados e círculos os quais indicam:

□ Nó de Decisão

○ Nó de Aleatoriedade

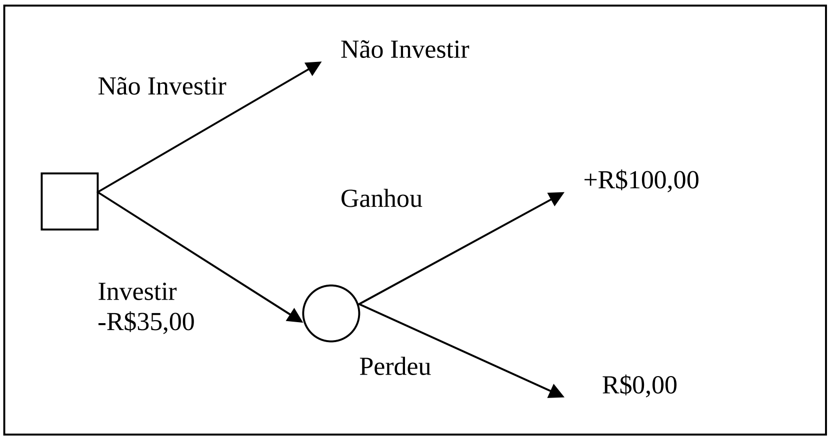
O leitor pode supor o seguinte exemplo: Você tem a oportunidade de ganhar R$100,00 se você acertar se a face de um dado é par ou ímpar. No entanto, essa oportunidade não é gratuita. Para jogar você deve pagar R$35,00. Há somente uma chance para investir. Você aceitaria?

<u>Como Avaliar essa Oportunidade?</u>

As típicas respostas são:

- Eu posso perder R$35,00, está barato esse jogo.
- Eu poderia ganhar R$100,00 e eu tenho muita sorte.
- Eu jogaria uma moeda para decidir.
- Eu preciso perguntar para a esposa.
- Eu não aposto em jogos.
- Minha taxa de retorno é .....

A decisão a ser tomada é investir ou não investir R$35,00 na *oportunidade* de receber R$100,00 ou R$0,00 dependendo do resultado do dado. A árvore de decisão será:



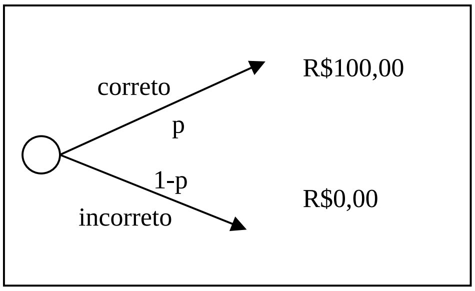
O que mais seria necessário para avaliar essa *oportunidade*?

A árvore de probabilidades ajuda a tomar a decisão para bons e maus resultados. A árvore incorpora um julgamento ao decisor sobre a probabilidade de sucesso e de sua chance complementar, que nesse caso significa a perda de investimento.

Que tipos de informações ajudariam ao decisor?

- O número de lados do dado.
- Conhecer a freqüência dos resultados do dado para ver sua honestidade.
- Quem jogará o dado?

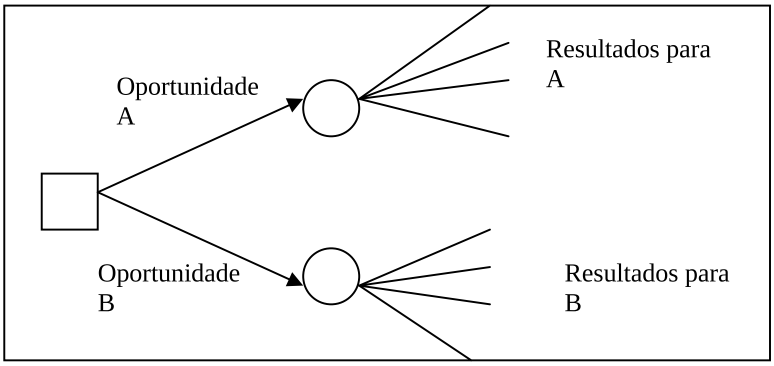
O ramo de aleatoriedade seria:



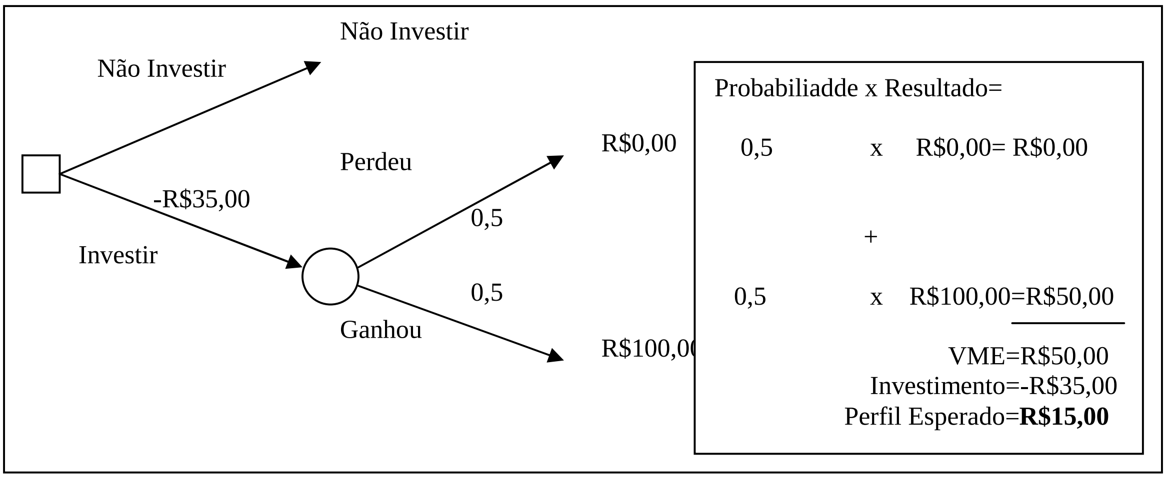
Existe uma importante distinção existente entre *oportunidade* e *resultados*.

Oportunidade é a soma dos possíveis *resultados*.

Isto é importante pois, nós podemos escolher somente oportunidades – não seus resultados!



Uma maneira de avaliar as decisões é quantificando-as através do ***valor monetário esperado*** (VME). O VME é a média ponderada dos resultados favoráveis e desfavoráveis ponderados pelas probabilidades. Para o exemplo do investidor que jogará o dado, a árvore de decisão seria:



Assim, neste caso, há uma boa chance de, jogando-se este jogo proposto o investidor receba como lucro o valor de R$15,00. O importante nesse caso, é perceber que existir uma chance de lucro não significa lucro! Existe uma probabilidade, de se puder jogar muitas vezes, o investidor receba em média por jogo um valor de R$15,00. A informação foi fundamental, pois por exemplo, sem saber se o dado tem 6 faces ou não, é impossível calcular as probabilidades, e o investidor estará jogando "às cegas". As probabilidades foram calculadas como:

$$p = \frac{3\,\text{pares}}{6\,\text{faces}} = 0{,}5$$

$$1 - p = 0{,}5$$

2.3 O Valor da Informação

Uma vez conhecendo-se a árvore de decisão é possível simular cenários para os eventos e analisar o que, por exemplo, uma mudança de parâmetros ocasionaria. Para tanto é necessário a aquisição de dados aleatórios ou de tabelas (em desuso) ou programas computacionais (tais como excel for windows, dentre outros).

Teorema de Bayes

O uso da informação como ferramenta para diminuir incertezas futuras é bastante utilizado na teoria da decisão. Essa informação aumenta as previsões sobre as probabilidades dos eventos, ocasionando assim melhor poder de decisão. O teorema de Bayes diz que : " Uma vez conhecida uma informação sobre um evento no passado, a probabilidade desse evento ocorrer no presente é uma ponderação entre as chances de repetição desse evento até o presente pela sua ocorrência no passado". Em termos matemáticos, é a divisão da probabilidade de intersecção entre dois eventos pela probabilidade da ocorrência do evento passado. Em fórmula seria,

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

para dois eventos A e B e será

$$P(A_i \mid B) = \frac{P(A_i) \times P(B \mid A_i)}{\sum_{i=1}^{n} P(A_i) \times P(B \mid A_i)}$$

para n eventos A_i mutuamente exclusivos.

Vamos observar sua aplicação no seguinte jogo:

Exemplo

Suponha que o jogo consiste em jogar uma moeda e observar sua face superior. Ganha primeiro quem tirar "coroa". No entanto, teremos 3 moedas para escolher uma, sendo que destas 2 são honestas e uma das moedas possui duas caras. Uma vez que um jogador jogou uma moeda, qual é a probabilidade de ter saído a moeda honesta se ao jogar se observou “cara” ?

Vamos nomear os seguintes eventos:

A_1={moeda honesta}
A_2={moeda de duas faces}
B={saiu cara}

Então, utilizando Bayes,

$$P(A_1 \mid B) = \frac{P(A_1) \times P(B \mid A_1)}{P(A_1) \times P(B \mid A_1) + P(A_2) \times P(B \mid A_2)}$$

Isto se interpreta assim: dado que saiu o evento B (saiu cara), qual a probabilidade de ter sido na moeda honesta (A_1)?

Para resolver esse problema, basta traduzir o problema para as probabilidades. Então nesse caso,

$$P(A_1)=2/3$$
$$P(A_2)=1/3$$

Uma vez que se tem duas honestas dentre 3 moedas no total e uma desonesta dentre 3 moedas no total. E ainda,

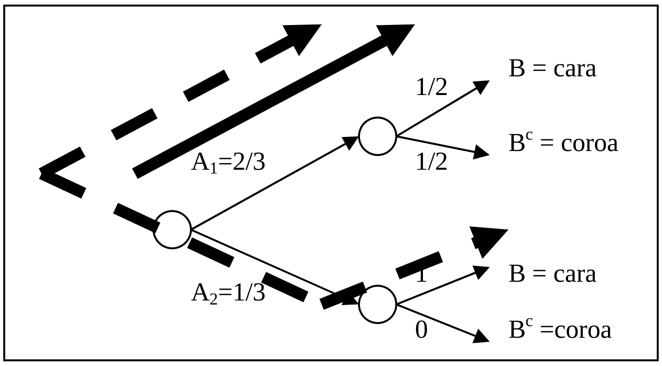
$$P(B|A_1)=1/2$$
$$P(B|A_2)=1$$

Sendo neste caso que a probabilidade de sair cara(B) na moeda honesta é ½ e a probabilidade de sair cara na desonesta é de 100% (A_2). Então substituindo na fórmula será:

$$P(A_1 \mid B) = \frac{(2/3) \times (1/2)}{(2/3) \times (1/2) + (1/3) \times 1} = \frac{1}{2} = 50\%$$

Ou seja, mesmo possuindo uma moeda desonesta, a presença de duas honestas fez com que as chances do jogo ser honesto (50% para cada) ficassem corretas!

Uma maneira mais simples de enxergar a teoria da informação e decisão Bayesiana é através da árvore de probabilidade como visto na seção anterior. Assim, no caso do jogo da moeda desonesta, a árvore seria:



Neste caso é só observar que a probabilidade será calculada dividindo o ramo indicado pela flecha contínua (produto das duas probabilidades), pela soma dos ramos indicados pelas flechas tracejadas.

Então o *valor da informação* agora pode ser somado aos valores previamente adquiridos pela estatística, tornando o cálculo de probabilidades para a tomada de decisão mais rico. Uma vez a empresa tendo em mãos dados sobre vendas, compras, patrimônios, ações, dívidas, enfim, qualquer tipo de variável, é possível através da estatística estimar o valor de cada ramo da árvore de decisão.

2.4 O Valor Monetário Esperado e a Volatilidade

Uma tomada de decisão empresarial, envolve sobretudo, oscilações de momento, de período de avaliação, de eventos econômicos, tanto indicando otimismo quanto pessimismo. Então, como saber se a decisão encontrada pela árvore é confiável e segura para avaliações de cenários. A saída é a utilização do *Método de Monte Carlo* que realiza através de repetições, cálculos com os números aleatórios, simulando como se a realidade estivesse alterando instântaneamente os cenários probabilísticos traçados.

Vamos Analisar o seguinte problema:

Aquisição de Empresas

Uma grande empresa deseja fazer uma aquisição de outras três empresas menores {A,B,C}, concorrentes do setor, em situação financeira complicada. A empresa compradora, no entanto, antes de realizar as aquisições, deseja traçar um cenário e avaliar a possibilidade de lucro nos meses após essas aquisições. As três empresas tem o seguinte comportamento:

Empresa A

Deseja lançar dois produtos novos. A empresa acredita que com 80% de chances, o produto A_1 implacará e fornecerá um lucro de R$10,00 por produto e com 20% de chances um prejuízo de R$1,00 por produto. Da mesma forma para o produto A_2 a empresa crê que com 40% de chances ele terá um lucro de R$4,00 por produto e 60% de chances de prejuízo de R$5,00 por produto. Essa empresa está em dúvida qual dos dois produtos ela lançará diante das concorrências do mercado e do setor. Existe por parte da diretoria 60% de chances de lançar o produto A_1 e 40% de chances de lançar A_2.

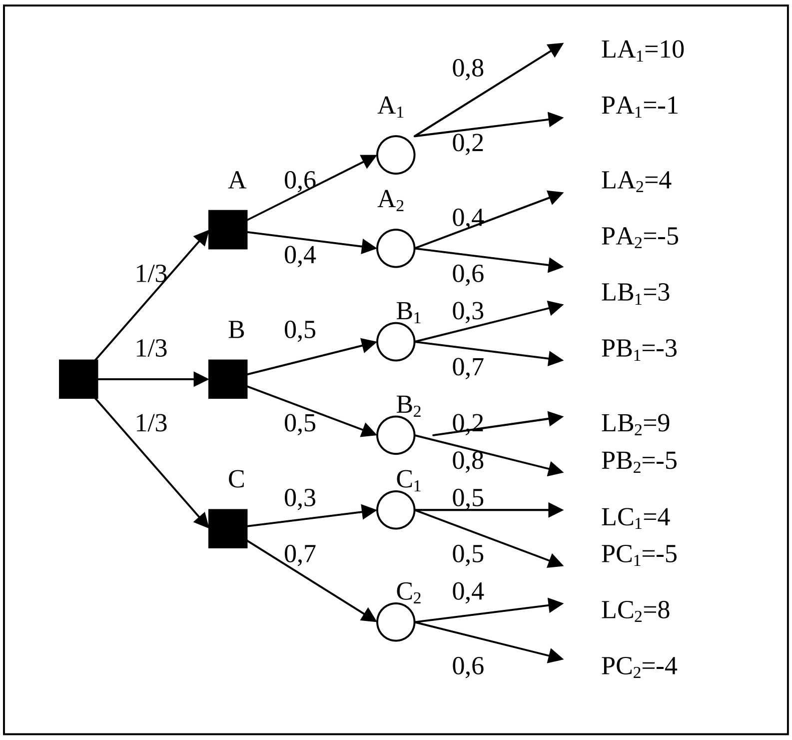
Empresa B

Deseja lançar dois produtos novos. A empresa acredita que com 30% de chances, o produto B_1 implacará e fornecerá um lucro de R$3,00 por produto e com 70% de chances um prejuízo de R$3,00 por produto. Da mesma forma para o produto B_2 a empresa crê que com 20% de chances ele terá um lucro de R$9,00 por produto e 80% de chances de prejuízo de R$5,00 por produto. Essa empresa também está em dúvida sobre qual dos dois produtos ela lançará diante das concorrências do mercado e do setor. Uma sondagem mostrou uma divisão entre os diretores sendo que 50% deseja lançar o produto B_1 e 50% de chances de lançar B_2.

Empresa C

Também deseja lançar dois produtos novos. A empresa acredita que com 50% de chances, o produto C_1 fornecerá um lucro de R$4,00 por produto e com 50% de chances um prejuízo de R$5,00 por produto. Da mesma forma para o produto C_2 a empresa crê que com 40% de chances ele terá um lucro de R$8,00 por produto e 60% de chances de prejuízo de R$4,00 por produto. Essa empresa também está em dúvida sobre qual dos dois produtos ela lançará diante das concorrências do mercado e do setor. Uma sondagem mostrou uma divisão entre os diretores sendo que 30% deseja lançar o produto C_1 e 70% de chances de lançar C_2.

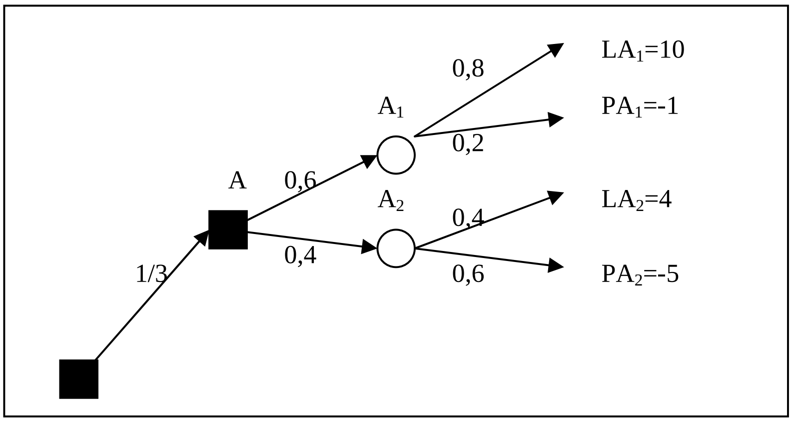
A árvore de decisão desse projeto será:



Vamos supor que a decisão da empresa que irá realizar a aquisição, seja a de primeiro adquirir a empresa com maior valor monetário esperado. Qual empresa deverá primeiro ser adquirida?

A análise através dos ramos fica bastante fácil. Partindo do quadrado representando uma decisão, deve-se multiplicar os valores dos ramos de trás-para-frente pelos respectivos lucros e prejuízos e então somá-los até o nó de decisão.

Valor Monetário Esperado - Empresa A



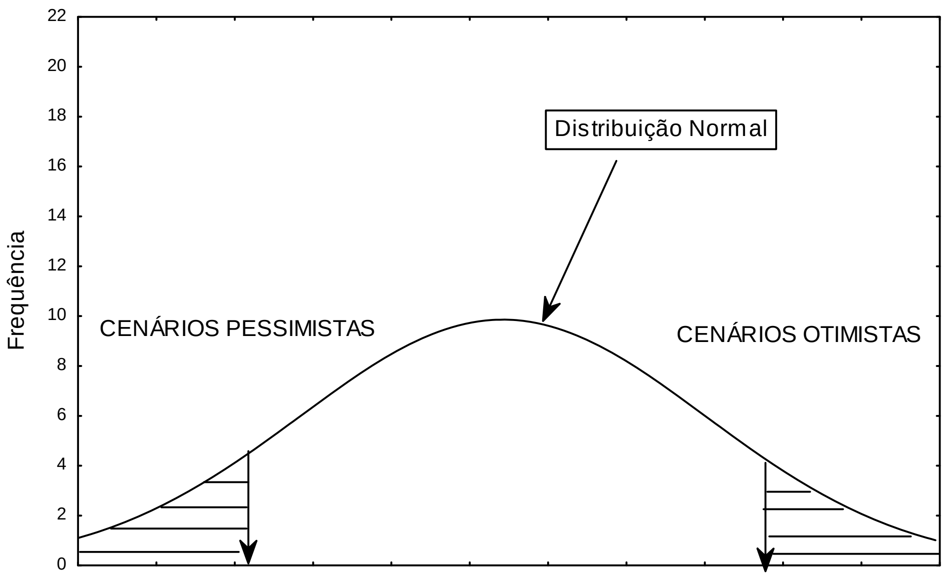
$$\text{VME(A)} = 0{,}6 \times \left(0{,}8 \times 10 + 0{,}2 \times (-1)\right) + 0{,}4 \times \left(0{,}4 \times 4 + 0{,}6 \times (-5)\right) = \text{R\$}4{,}12$$

Com os mesmos tipos de cálculos os valores para as empresas B e C serão:

$$\text{VME(B)} = 0{,}5 \times \left(0{,}3 \times 3 + 0{,}7 \times (-3)\right) + 0{,}5 \times \left(0{,}2 \times 9 + 0{,}8 \times (-5)\right) = -\text{R\$}1{,}7$$
$$\text{VME(C)} = 0{,}3 \times \left(0{,}5 \times 4 + 0{,}5 \times (-5)\right) + 0{,}7 \times \left(0{,}4 \times 8 + 0{,}6 \times (-4)\right) = \text{R\$}0{,}41$$

Neste caso, a decisão seria primeiro fazer aquisição da empresa A, depois a empresa C e por último a empresa B. No entanto, deve-se levar em conta, que como valores estimados, existe por trás uma volatilidade que deve ser adicionada aos valores esperados. A adição e subtração dessa volatilidade aos valores monetários esperados, cria um intervalo, conhecido como intervalo de confiança para uma estimação estatística. O termo volatilidade é mais conhecido em estatística como variância, cuja extração da raiz quadrada fornecerá o desvio-padrão. Sabe-se da teoria de inferência, que a soma e subtração de um desvio padrão ao valor monetário esperado, fornecerá às estimativas uma confiabilidade de 68%. Isso corresponde a área em baixo da curva normal de probabilidades. Se quisermos mais confiança devemos trabalhar com dois desvios padrões para cima e para baixo da média, ou seja,

Média $\pm$ 2 x Desvio Padrão .... Confiança de 95% nas previsões
Média $\pm$ 3 x Desvio Padrão..............................Confiança de 99% nas previsões



O gráfico da figura anterior apresenta como cenário *pessimista* a área da curva normal atrás do valor monetário esperado e como *otimista* a área da curva normal acima do valor monetário esperado. Essa ilustração pode ser invertida, dependendo do contexto do que se entende por otimista e pessimista.

A Volatilidade (desvio-padrão) de Cenários

O Cálculo do desvio-padrão que ajudará a criar o intervalo de confiança é bastante simples. A fórmula é:

$$dp = \pm\sqrt{VME(x^2) - VME(x)^2}$$

Para o caso das empresas, primeiro deve-se calcular o valor monetário esperado dos valores quadráticos, ou seja,

$$VME(A^2) = 0{,}6\times\left(0{,}8\times10^2 + 0{,}2\times(-1)^2\right) + 0{,}4\times\left(0{,}4\times4^2 + 0{,}6\times(-5)^2\right) = R\$56{,}68$$
$$VME(B^2) = 0{,}5\times\left(0{,}3\times3^2 + 0{,}7\times(-3)^2\right) + 0{,}5\times\left(0{,}2\times9^2 + 0{,}8\times(-5)^2\right) = R\$22{,}6$$
$$VME(C^2) = 0{,}3\times\left(0{,}5\times4^2 + 0{,}5\times(-5)^2\right) + 0{,}7\times\left(0{,}4\times8^2 + 0{,}6\times(-4)^2\right) = R\$30{,}79$$

Então, por exemplo para 68% de confiança nas estimativas, os desvios padrões dos valores monetários esperados nas empresas serão:

$$dp(A) = \pm\sqrt{VME(A^2) - VME(A)^2} = \pm\sqrt{56{,}68 - (4{,}12)^2} = \pm\sqrt{39{,}7} = \pm R\$6{,}3$$
$$dp(B) = \pm\sqrt{VME(B^2) - VME(B)^2} = \pm\sqrt{22{,}6 - (1{,}7)^2} = \pm\sqrt{19{,}7} = \pm R\$4{,}4$$
$$dp(C) = \pm\sqrt{VME(C^2) - VME(C)^2} = \pm\sqrt{30{,}79 - (0{,}41)^2} = \pm\sqrt{30{,}62} = \pm R\$5{,}5$$

Então os cenários otimistas e pessimistas para as 3 empresas serão:

EMPRESAS	PESSIMISTA	OTIMISTA
A	-R$2,18	R$10,42
B	-R$6,10	R$2,70
C	-R$5,09	R$5,91

Estes intervalos fornecerão estimativas mais seguras para a empresa que deseja fazer aquisição, uma vez que, estatisticamente poderá ocorrer que em algum mês os lucros estimados não sigam as ordens colocadas pelos valores monetários estimados. Ou seja, poderá existir algum mês, onde por exemplo, a empresa B fornecerá menos prejuízo que a empresa A.

2.5 Simulando Modelos com Incertezas baseadas nas Distribuições de Probabilidades

O primeiro passo antes de inserir aleatoriedade nos modelos é a realização de uma estatística descritiva como a apresentada no tópico 1, para um ajuste de qual distribuição de probabilidades regem as flutuações das observações. Na hora de inserir um número aleatório ele deve satisfazer um *gerador de números aleatórios* que obedeça a distribuição normal.

Gerando números com distribuição de probabilidade uniforme

A regra para gerar números com distribuição de probabilidades uniforme em um intervalo [a,b] será:

$$x = a + (b - a)\times \text{aleatório}$$

onde “x” é o número que respeita a distribuição uniforme e *aleatório* é o número gerado de maneira aleatória usando por exemplo a função ALEATÓRIO( ) do Excel.

Gerando números com distribuição de probabilidade normal

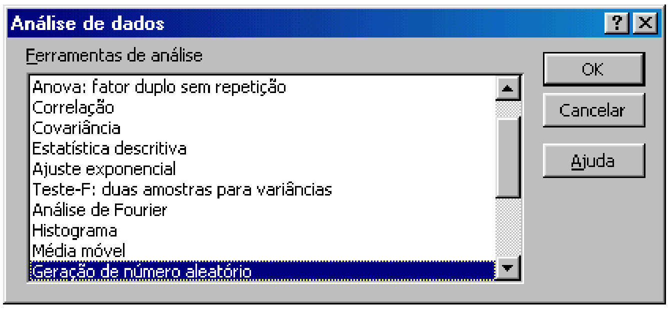
Neste caso, para gerar números com distribuição normal necessita-se de números gerados por outros geradores. É necessário primeiramente que sejam somados 12 números com distribuição uniforme. O resultado desse total é subtraído de 6, multiplicado pelo desvio padrão desejado para a distribuição dos números somados com a média desejada para a geração. Em termos matemáticos,

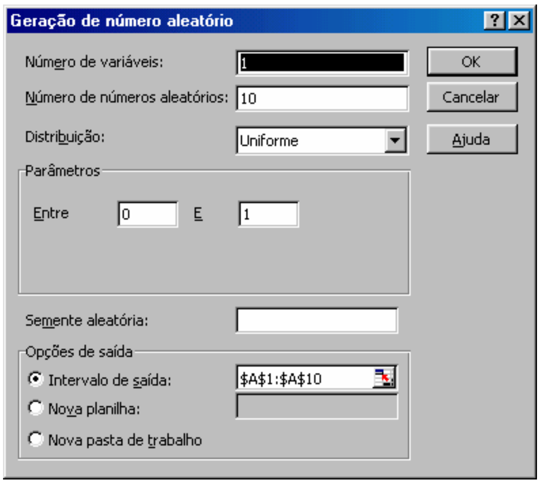
$$x = (\text{total} - 6) \times s + \text{média}$$

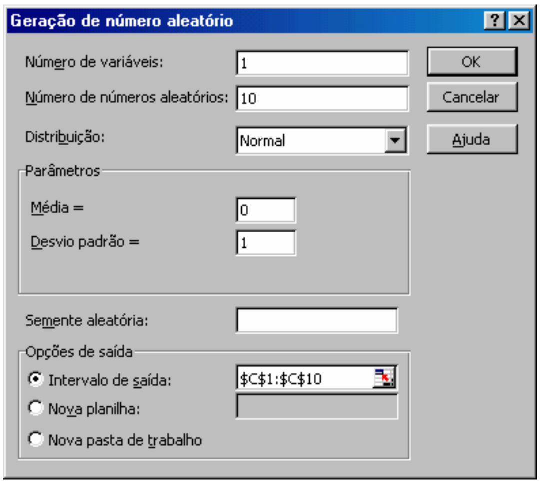
onde "total" é a soma de 12 números com distribuição normal, "média" é a média desejada para os números e "s" o desvio padrão desejado para esses números.

Os Geradores do Excel

Apesar das fórmulas de geração serem apresentadas no ítem anterior, todas já estão programadas em diversos *softwares*, tais como no Excel. Isso é bastante interessante, uma vez que simulações poderão ser realizadas na planilha usando a distribuição das flutuações mais próximas da distribuição real dos erros de amostragens.







As figuras anteriores mostram as maneiras de acessar os geradores do Excel. Os passos são: ao abrir o Excel ir até o menu *ferramentas*, depois *análise de dados* e então escolher *geração de número aleatório*, conforme mostrado nas telas anteriores. Dentro do gerador existem várias opções de escolha de distribuição de probabilidades. Na figura da distribuição uniforme a figura mostra a escolha de um intervalo entre 0 e 10 para a geração de 10 números. Observa-se que nesta janela o intervalo [a,b] da distribuição uniforme deve ser preenchido, que aparece como padrão 0 e 1.

A distribuição normal é a mais utilizada, graças ao teorema do limite central da probabilidade e estatística. Essa distribuição está também programada no Excel. Os parâmetros a serem preenchidos são a média e o desvio padrão.

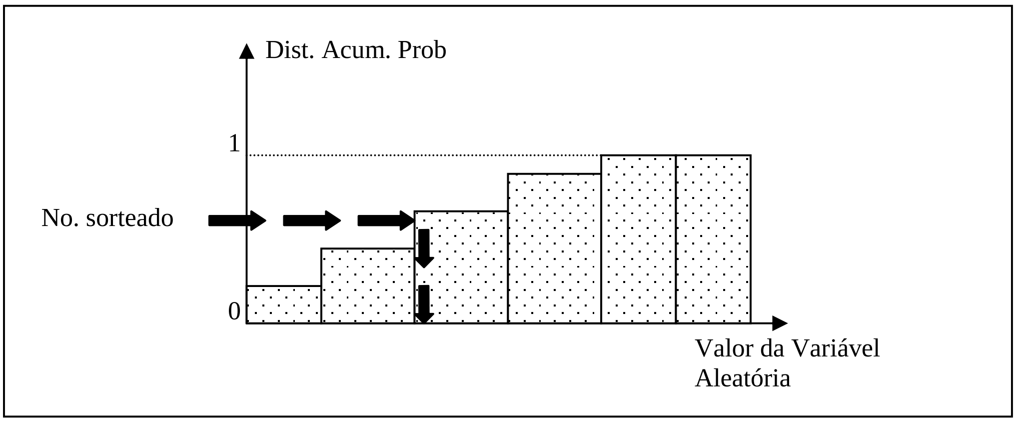
2.6 O Método de Monte Carlo e a Tomada de Decisão

O método de Monte Carlo é bastante conhecido pela sua ampla aplicação e diversificação de áreas em que ele auxiliar na criação de cenários econômicos. O método consiste em utilizar números aleatórios gerados através de alguma distribuição de probabilidade e exaustivamente os substituir nos modelos para alteração do padrão determinístico. Consiste em simular as flutuações em torno de valores médios observados.

Com isso é possível, alterando-se os diversos tipos de modelos, criar situações e fazer avaliações sobre os resultados das simulações. Quanto mais simulações forem realizadas, mais certeza e confiabilidade serão obtidos nos resultados.

Os passos do algoritmo de Monte Carlo consistem em:

- Determinar a estatística com a freqüência dos valores observados para cada variável.
- Criar a distribuição acumulada das probabilidades.
- Sortear um número aleatório uniforme.
- Avalia-se o valor na abcissa da distribuição acumulada.



Suponhamos que para o problema de tomada de decisão, das empresas A, B e C tenhamos a seguinte tabela de números aleatórios.

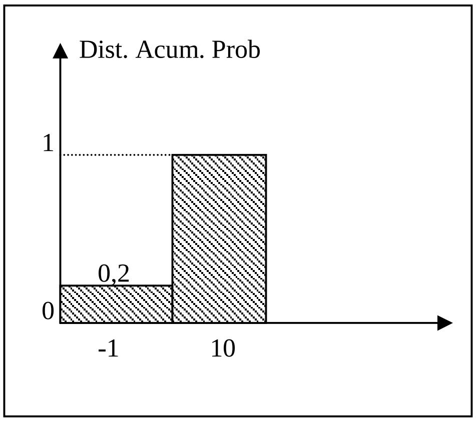
0,382	0,731223	0,001251
0,100681	0,065188	0,563585
0,596484	0,777856	0,193304
0,899106	0,796442	0,808741
0,88461	0,870022	0,585009
0,958464	0,963958	0,479873
0,014496	0,130253	0,350291
0,407422	0,610736	0,895962

0,863247	0,75988	0,82284
0,138585	0,454299	0,746605

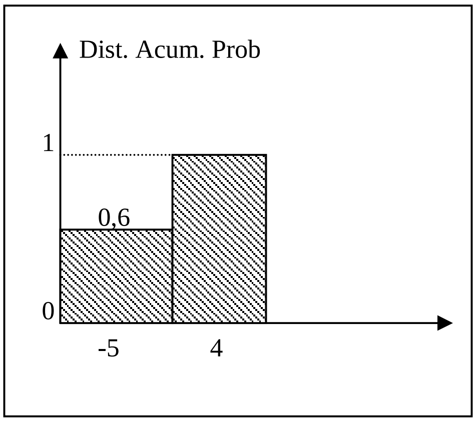
Esses números servirão como base para uma simulação sobre os ramos da árvore de decisão. Antes porém, devemos criar as distribuições acumuladas, que servirão como base para a escolha dos valores. Suponhamos que se deseja a simulação dos lucros das três empresas num mês, L(A), L(B) e L(C).

<u>Produto A1</u>

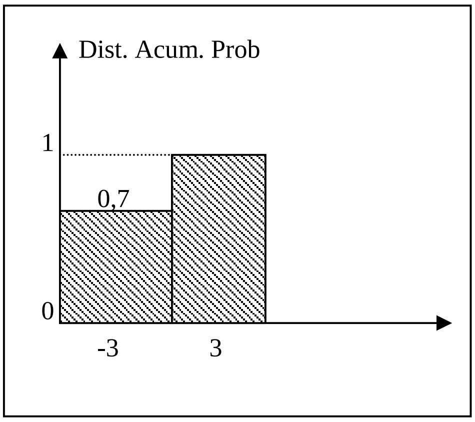
A estatística desse produto, produzirá a seguinte distribuição acumulada:



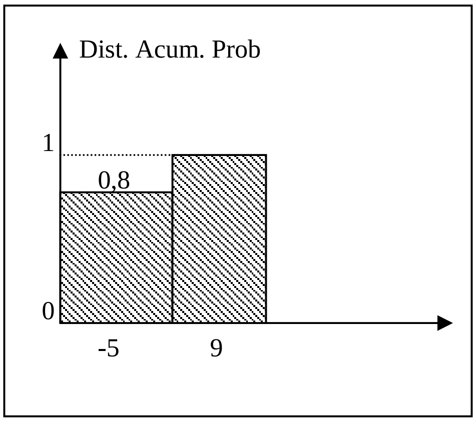
<u>Produto A2</u>



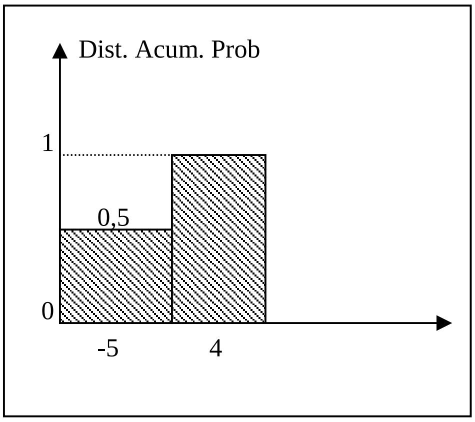
<u>Produto B1</u>



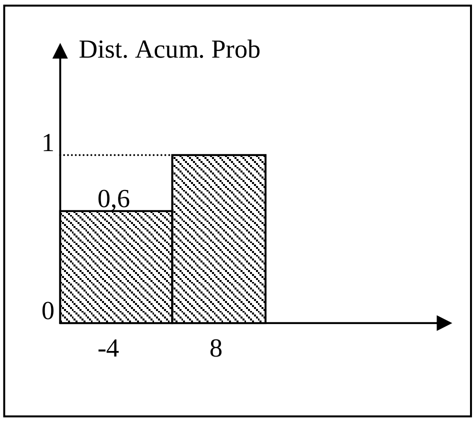
<u>Produto B2</u>



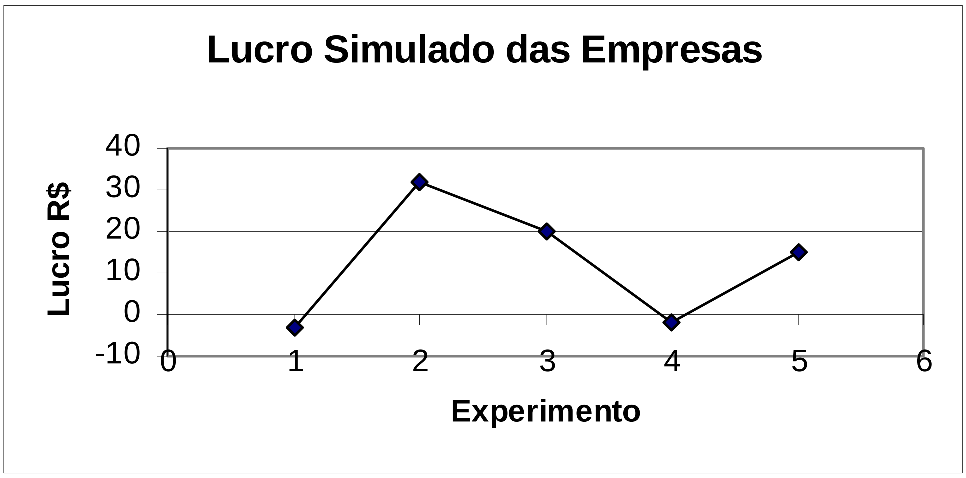
<u>Produto C1</u>



Produto C2



Então, com essas estatísticas dos ramos para os lucros das empresas e dos histogramas apresentados anteriormente, é possível fazer uma simulação dos lucros esperados. Para o produto A1, o primeiro número aleatório (da tabela de números aleatórios) seria 0,382, maior portanto que 0,2 (valor do prejuízo de –R$1,00) do histograma acumulado. Assim nesse passo, o produto A1 forneceu um lucro de R$10,00. O segundo número aleatório, 0,73, é maior que 0,6, valor do histograma acumulado de A2. Nesse passo, A2 forneceria um lucro de R$4,00. Adicionado ao lucro do produto A1, a empresa A nesse experimento forneceria um lucro de (L(A1)+L(A2)) R$14,00. O gráfico e a tabela a seguir fornecem todos os lucros simulados nas empresas utilizando a tabela de números aleatórios.



Nº EXPERIMENTO	PRODUTO	Nº SORTEADO	LUCRO PRODUTO	LUCRO EMPRESA
1	A1	0,382	10	**-3**
	A2	0,731	4	
	B1	0,0012	-3	
	B2	0,1006	-5	
	C1	0,0651	-5	
	C2	0,5638	-4	
2	A1	0,596	10	**32**
	A2	0,777	4	
	B1	0,193	-3	
	B2	0,899	9	
	C1	0,796	4	
	C2	0,808	8	
3	A1	0,884	10	**20**
	A2	0,870	4	
	B1	0,585	-3	
	B2	0,958	9	
	C1	0,963	4	
	C2	0,479	-4	
4	A1	0,014	-1	**-2**
	A2	0,130	-5	
	B1	0,350	-3	
	B2	0,407	-5	
	C1	0,610	4	
	C2	0,895	8	
5	A1	0,863	10	**15**
	A2	0,759	4	
	B1	0,822	3	
	B2	0,138	-5	
	C1	0,454	-5	
	C2	0,746	8	
	LUCRO TOTAL ESPERADO			**R$ 12,40**

É claro que 5 simulações não fornecem idéia da variação dos valores. Uma simulação de Monte Carlo só se torna útil para mais de 1000 repetições. No entanto, como exemplo didático, pode-se perceber que a aquisição das 3 empresas A, B e C trará um lucro médio de cerca de R$10,00 por produto vendido após as aquisições.

TÓPICO 3
Método Estatístico de Depreciação

3.1 Introdução

Para a maioria das grandes empresas, o primeiro passo relativamente ao processo de realização de investimentos consiste na preparação de um *orçamento de investimentos* anual, que é uma lista dos projetos de investimento e um desdobramento das despesas de investimentos por fábricas e divisões. Uma das maneiras de avaliar uma performance operacional de uma empresa é medindo a taxa de rentabilidade esperada. É demasiado fácil calcular a taxa de rentabilidade verdadeira ou econômica para uma ação ordinária que é continuamente transacionada. Apenas deve-se registrar os recebimentos (dividendos) para o ano, somar a alteração de preço durante o ano e dividir pelo preço inicial:

$$\text{taxa de rentabilidade} = \frac{\text{recebimentos} + \text{alteração de preço}}{\text{preço inicial}} = \frac{C_1 + (P_1 - P_0)}{P_0}$$

O numerador da expressão da taxa de rentabilidade é denominado *resultado econômico*:

Resultado econômico = fluxo de tesouraria + alteração do valor presente

Qualquer redução do valor presente representa uma depreciação econômica; qualquer aumento do valor presente representa uma depreciação econômica negativa. Portanto,

Depreciação Econômica = redução do valor presente

e

Resultado Econômico = fluxo de tesouraria – depreciação econômica

Uma dificuldade em quantificar o resultado econômico e a rentabilidade é a de calcular o valor presente. Pode-se observar o valor de mercado das ações em carteira, se estas forem ativamente negociadas, mas poucas unidades industriais, divisões ou projetos de investimento tem suas próprias ações transacionadas no mercado de ações. Pode-se observar o valor presente de mercado de todos os ativos de uma empresa, mas não o de cada um individualmente. Os contabilistas geralmente nos fornecem o valor contábil líquido, que consiste no custo inicial menos as amortizações, calculadas de acordo com critérios arbitrários. Muitas empresas utilizam o valor contábil para calcular a rentabilidade contábil do investimento (RCI):

$$RCI = \frac{\text{fluxo de tesouraria - amortização contábil}}{\text{valor contábil inicial}}$$

ou

$$RCI = \frac{C_1 + \left(VC_1 - VC_0\right)}{VC_0}$$

No entanto, se a amortização contábil e a depreciação econômica forem diferentes (raramente são idênticas), então os indicadores de rentabilidade contábil estarão errados, ou seja, não quantificarão a verdadeira rentabilidade. Para essa correção se faz necessário diversos tipos de estimativas sobre o valor de depreciação real.

3.2 Modelos de Depreciação

O conceito de depreciação é de uso freqüente e importante. Significa a perda progressiva da eficiência funcional de bens imobiliários como edifícios, máquinas, instalações, veículos, etc. Para a contabilização desse valor de perda de valor, existem inúmeros métodos para se realizar uma estimativa de depreciação de valores empresariais. O grande problema é que, sendo estes métodos quantitativos, podem servir para os objetivos de uma empresa ou governo, mas não para outros. Os diversos modelos de depreciação podem ser classificados em três grupos:

- Modelos em que as cotas de depreciação são mantidas constantes ao longo da vida. Um modelo nesse caso é o conhecido como **depreciação linear.**
- Modelos que calculam quotas de depreciação maiores no início da vida útil do empreendimento. Estes modelos são conhecidos como modelos de **depreciação acelerada**. Fazem parte o *modelo dos dígitos dos anos* (ou de Cole) e o *modelo do saldo decrescente* (ou de Matheson).
- Modelos de depreciação lenta que calculam cotas menores no início da vida útil do bem ou empreendimento. O *modelo do fundo de amortização* faz parte desse grupo.

<u>Modelo da Depreciação Linear</u>

É o modelo clássico mais simples de depreciação. É o único reconhecido pela legislação do Imposto de Renda. Sua fórmula é:

$$c_k = c - \frac{k \times c}{n}$$

onde

c: custo de aquisição

c_k: valor contábil após o decurso de k períodos de vida útil

n: número de períodos de vida útil.

k: número de períodos de vida útil decorridos, com $0 \le k \le n$.

Caso o investimento tenha um valor residual (r), o modelo se transforma em

$$\boxed{c_k = c - \frac{k \times (c - r)}{n}}$$

A taxa de **depreciação** neste modelo será

$$\boxed{d = \frac{c - r}{n}}$$

Modelo da Soma dos Dígitos dos Anos (Cole)

Este é um modelo onde a depreciação é acelerada, isto é, a depreciação é maior no início do período do que no final da vida útil. Neste modelo a quotas de depreciação nada mais são do que o valor depreciável (c – r) que obedece a seguinte lei de formação:

$$\frac{n-k+1}{1+2+3+\cdots+k+\cdots+n} \quad 1 \le k \le n$$

Como o denominador é a soma dos n primeiros naturais, pela soma de uma progressão aritmética

$$1+2+...+n = \frac{n\times(n+1)}{2}$$

Então o modelo para o valor contábil no final do período k é:

$$\boxed{c_k = c - \frac{2\times k\times(c-r)\times(n-k+1)}{n\times(n+1)}}$$

A taxa de **depreciação** será:

$$\boxed{d_k = \frac{2\times(c-r)\times(n-k+1)}{n\times(n+1)}}$$

Modelo do Saldo Decrescente (Matheson)

O modelo assume que o valor do bem decresce a uma taxa percentual constante que segue a fórmula:

$$p = 1 - \sqrt[n]{\frac{r}{c}}$$

Então o valor contábil no período k+1 será:

$$\boxed{c_k = c_{k-1} - p\times c_{k-1}}$$

E a depreciação será

$$\boxed{d_k = p\times c_{k-1}}$$

Se a taxa de valor residual é nula (r = 0) a fórmula para “p” não tem significado. O que se faz então é estimar uma taxa percentual que seja adequada ao investimento.

3.3 Experimento Real de Depreciação

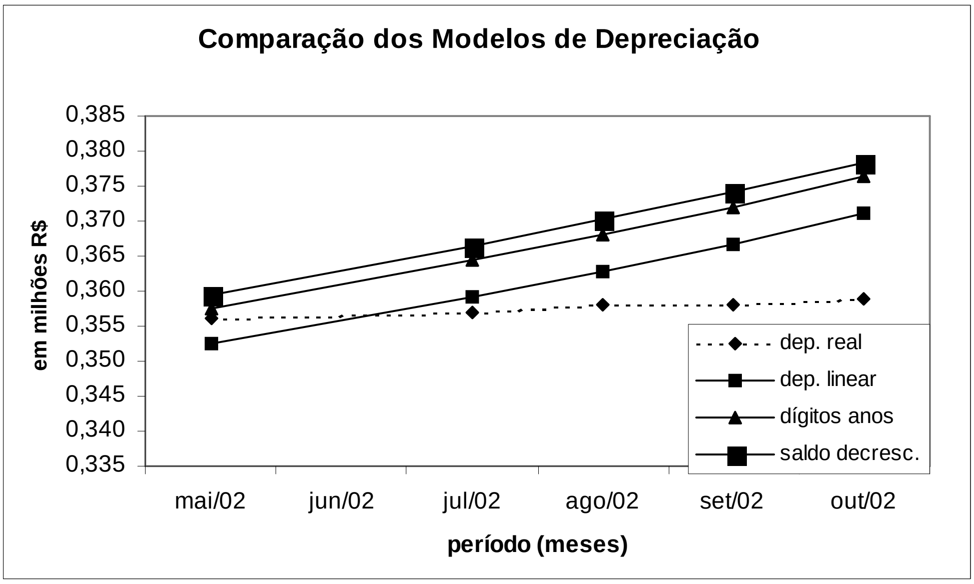
Os modelos anteriores, são modelos clássicos e bastante difundidos na área de finanças. No entanto, como a depreciação é bastante subjetiva em termos práticos, esses modelos podem levar a super-depreciação de um ativo ou sub-depreciação, sendo nos dois casos um problema sério. Todos os métodos, no entanto, para fim da legislação de imposto de renda, servem como parâmetros para abatimentos. Dentre todos os modelos, o único aceito pela legislação do imposto de renda no Brasil, é o método da depreciação linear. Esses métodos estão todos programados no *microsoft excel* na biblioteca de fórmulas de finanças:

- Modelo de Depreciação Linear – Função Financeira **DPD**.
- Modelo da Soma dos Dígitos dos Anos – Função Financeira **SDA**.
- Modelo do Saldo Decrescente – Função Financeira **BD**.

Um exemplo de uma empresa foi utilizado para a demonstração dos erros desses métodos de depreciação e sua perda de poder de previsão ao longo do tempo. A empresa apresentou em seu relatório anual o valor do investimento e a depreciação dos ativos permanentes.

2002	**Depreciação (milhões R$)**	**Investimento (milhões R$)**	**Dep. Linear**	**Dig. Anos**	**Saldo Decresc.**
Maio	0,356	23,971	0,3525	0,358	0,360
Julho	0,357	24,429	0,3593	0,364	0,366
Agosto	0,358	24,678	0,3629	0,368	0,370
Setembro	0,358	24,938	0,3667	0,372	0,374
Outubro	0,359	25,228	0,3710	0,376	0,378

Para os três modelos foi realizado um ajuste com a finalidade de comparação da previsão com o valor real de depreciação. Para a dependência linear, utilizou-se r = 0 e vida útil de n = 68 meses. Para o método dos dígitos dos anos, r = 0 e o período transcorridos de depreciação de k = 34 meses. Para o método do saldo decrescente, faz-se necessário ajustar a percentagem “p”. Então supôs-se que o valor residual era r = R$ 8,8 (milhões), para vida útil de n = 68 meses e k = 34 meses, como os outros métodos. O resultado é apresentado na tabela anterior e no gráfico a seguir.



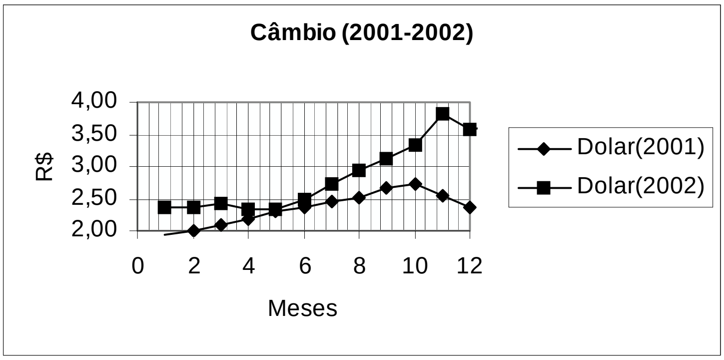
Percebe-se claramente que os métodos clássicos possuem uma tendência em depreciar continuamente os ativos, não refletindo para mais de dois períodos a realidade de depreciação. Mesmo para um único período, os modelos ou super-depreciam ou sub-depreciam. Por que esses modelos erram?

O fator preponderante, é que todos esses modelos seguem uma tendência linear de previsão e mesmo que investimentos ou políticas de investimentos se alterem durante o período fiscal, os modelos não tem como prever essas alterações.

Assim, torna-se importante trabalhar e enfocar modelos que utilizam informações passadas como uma realimentação do poder de estimação.

3.3 Modelos de Séries Temporais

Em muitas áreas de conhecimento as observações de interesse são obtidas em instantes sucessivos de tempo. Por exemplo, a cada segundo, ações são compradas ou vendidas pelo sistema eletrônico da BOVESPA. Chama-se de *série temporal* um conjunto de observações ordenadas no tempo.



Um mesmo fenômeno pode ser medido por dois períodos (ou por dois órgãos diferentes) e apresentar características muito diferentes. A figura anterior apresenta dois gráficos para a taxa de câmbio do dólar no Brasil em dois anos consecutivos. Esses gráficos são conhecidos como trajetórias do processo analisado (dólar). Cada trajetória é conhecida como uma *série temporal* e aqui será representada por x(t). Assim um fenômeno pode ser observado em diversos tempos $x(t_1)$, $x(t_2)$,...,$x(t_n)$, compondo uma série temporal e essa série pode ser analisada tendo como objetivos principais:

- Modelagem do Processo sob consideração.
- Obtenção de Conclusões em termos Estatísticos.
- Avaliação do modelo pressuposto para previsão do processo.

Um modelo para séries temporais bastante utilizado nas literaturas específicas supõe que a série $x(t_1)$, $x(t_2)$,...,$x(t_n)$ pode ser escrita como a soma de três componentes: uma tendência T(t), uma sazonal S(t) e uma componente aleatória w(t):

$$x(t) = T(t) + S(t) + \omega(t)$$

A componente de tendência T(t) em séries econômicas é causada por fatores que são medidos durante períodos longos de tempo. Ela possui um movimento suave durante o tempo de acompanhamento e indica um crescimento ou decréscimo dos valores observados para o intervalo de tempo estudado. A componente sazonal S(t) surge quando as observações são realizadas dentro de um subperíodo de amostragem, por exemplo: Se o estudo for anual, a sazonalidade poderá aparecer mensalmente, trimestralmente ou semestralmente. Ela é uma variação periódica de curto período. Já a componente w(t) é concebida como uma perturbação puramente aleatória pertencente ao aparato de coleta de dados mesmo como uma representação de interações desconhecidas no modelo adotado. Essa componente deverá possuir média nula e variância constante e é conhecida como ruído branco.

3.4 Estimação da Tendência de Valores Econômicos

Para o presente texto, será considerado o modelo sem a componente sazonal S(t). Embora ela seja muito importante, uma vez que se entende o funcionamento da tendência a adição dos termos S(t) e seus métodos particulares de estimação poderão ser entendidos sem problemas maiores. Portanto, o modelo adotado para uma determinada série temporal será:

$$x(t) = T(t) + w(t)$$

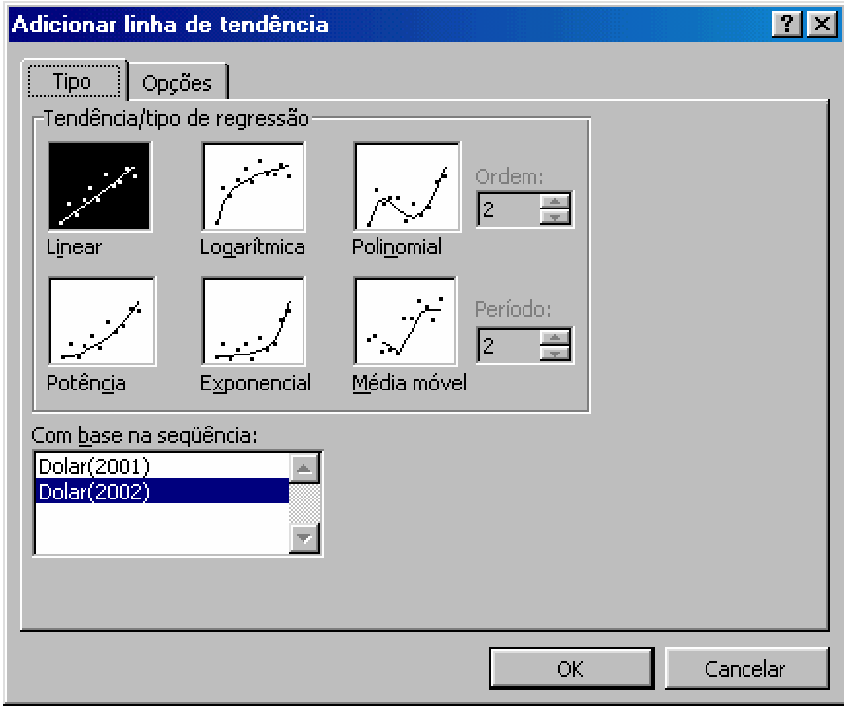
Um procedimento muito comum para a estimativa de T(t) é ajustar uma curva aos valores coletados (dados reais conhecidos). Tradicionalmente são utilizadas as seguintes funções:

- Polinômios: $T(t) = b_0 + b_1 t + b_2 t^2 + \cdots + b_n t^n$
- Exponenciais: $T(t) = ae^{bt}$
- Logísticas: $T(t) = \log(a) + bt$

O problema passa a ser então a estimação dos parâmetros “a” e “b” relacionados as funções. Com a adição de computadores, os métodos estatísticos utilizados para essa estimação de parâmetros são partes integrantes de diversos programas computacionais. No *microsoft excel*, por exemplo pode-se seguir os seguintes passos:

- Clicar com o botão direito do mouse em cima do conjunto de dados plotados em gráfico.
- Escolher “Adicionar Linha de Tendência”
- Escolher o tipo de tendência que melhor caracterize o conjunto de valores observados dentre os existentes: linear, logarítma, polinomial, potência, exponencial e média móvel.

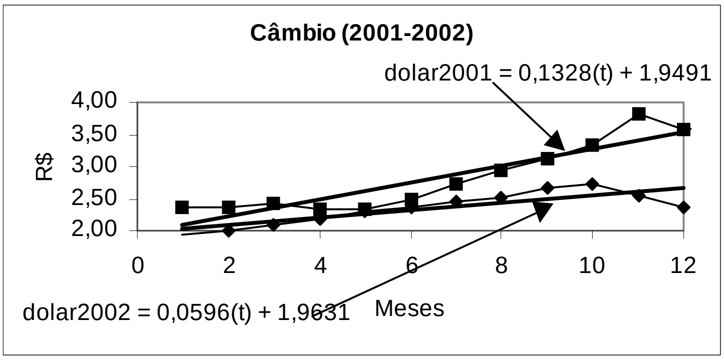
Os tipos estão apresentados na tela a seguir do Excel. Uma vez escolhido o tipo de tendência mais característica para os dados, pode-se ainda formatar se o modelo será linear ou não, modificando a ordem dos termos (quadrado, cubo, etc.) ao lado das tendências.



Para o exemplo do dólar, se for ajustada uma tendência linear, o modelo para da série sazonal para os dois anos seria:

- Para 2001: $x(t) = 0{,}1328 \times t + 1{,}9491 + w(t)$
- Para 2002: $x(t) = 0{,}0596 \times t + 1{,}9631 + w(t)$

O resultado dessa estimativa é mostrado comparativamente na figura a seguir.



Como a tendência escolhida foi linear, o gráfico obviamente é uma reta com as equações já descritas. Cabe lembrar que w(t) é um número aleatório com média nula e variância finita. Então os modelos estão prontos para o uso do método de Monte Carlo na realização de simulações e criações de cenários, como já descritos anteriormente. Esse ruído w(t) deve ser normalmente distribuído e essas simulações também podem ser realizadas no excel, como as já apresentadas no tópico de criações de cenários.

3.5 Modelo de Depreciação usando Séries Temporais

Como visto nas seções 3.2 e 3.3, os modelos tradicionais de estimação de depreciação não levam em conta informações novas para se auto corrigirem, ou mesmo informações passadas para melhorar as estimativas de tendências de depreciação. Isso torna, muitas vezes uma depreciação superestimada, como visto no exemplo da seção 3.3. Vamos denotar I(t) como o montante de investimento e D(t) o montante de depreciação no período t. Cada investimento é assumido ter uma vida útil de T períodos. No caso determinístico, tem-se a seguinte condição satisfeita:

$$D(t) = I(t) \text{ para todo } t \geq T$$

Mas no caso estocástico (perturbação aleatórias), o fluxo de caixa somente será correspondente a condição anterior se o operador de média for utilizado. Um investimento no período t denotado por I(t), é distribuído dentro de uma população de oportunidades com média desconhecida $\bar{I}(t)$. Esse investimento pode ser representado como

$$I(t) = \bar{I}(t) + w(t)$$

onde w(t) tem média zero e variância constante σ^2. Então, para o caso estocástico, a relação entre depreciação e investimento altera-se para:

$$E[D(t)] = \bar{I}(t) \text{ para todo } t \geq T$$

Em outras palavras, no caso estocástico, a depreciação média é que é equivalente ao investimento médio no final do período de vida útil, e essa depreciação flutuará de acordo com a característica dos investimentos (constante, sazonal, esporádicos, etc.). Se for adotado que o valor recuperado é nulo, alternativamente, I(t) pode ser visualizado como um conjunto de valores recuperados (ou salvos). Uma firma pode escolher uma tabela de depreciação (d_1, d_2, d_3,..., d_T) onde $d_i \geq 0$, e i = 1,2,3,...,T. Cada investimento I(t) é então depreciado num montante $d_1I(t)$ em seu primeiro ano de existência, $d_2I(t)$ no segundo ano, até o final do período de vida útil T. A depreciação total depois de T períodos será:

$$\sum_{i=1}^{T} d_i = 1$$

O montante de depreciação num período t (t ≥T) será denotado por uma série temporal seguindo o modelo de tendência somado ao ruído aleatório w(t):

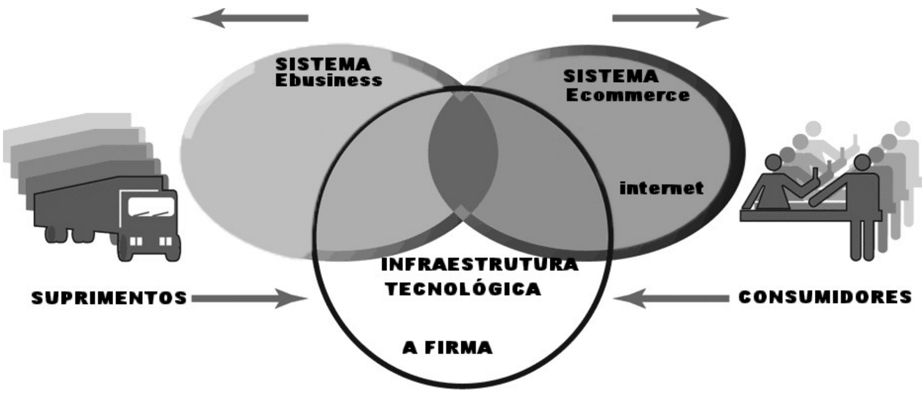
$$D(t) = d_1 I(t) + d_2 I(t-1) + \cdots + d_T I_{T+1} + w(t)$$

Ou seja, o montante de depreciação será uma transformação linear dos investimentos ao longo do período T. No caso de eventos aleatórios (processos estocásticos) faz-se necessários estimadores não tendenciosos para os termos d_i. O que se busca então, é a melhor tabela para a depreciação d_i de forma a tornar o montante de depreciação D(t) com a menor variância possível nos cenários econômicos.

TÓPICO 4
Projeção de Valor Econômico com Métodos Estocásticos – Filtro de Kalman

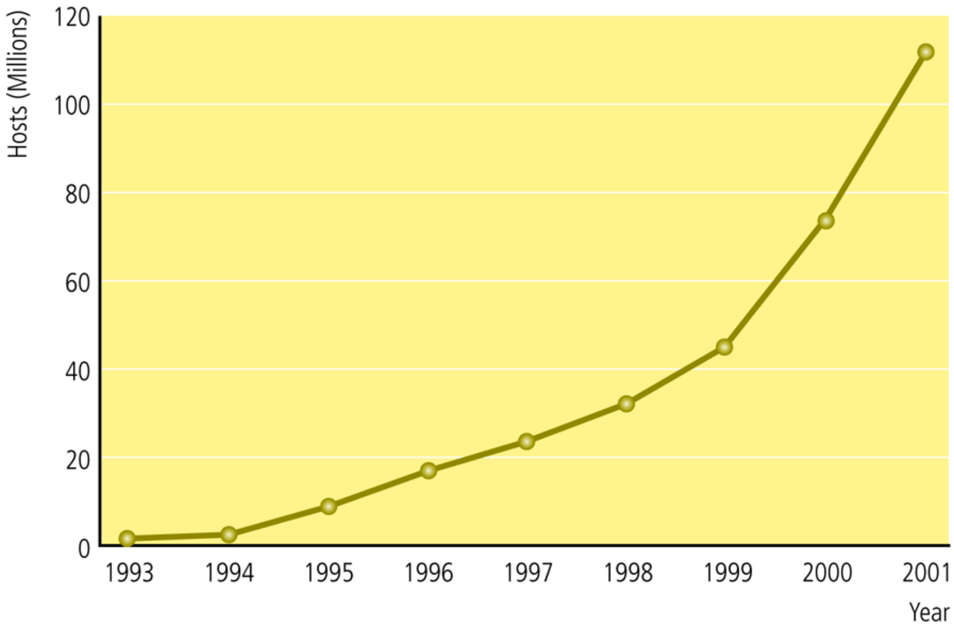
4.1 Introdução

Nos tópicos anteriores foram apresentadas técnicas clássicas para apresentação de dados e estimação de valores econômicos, sobretudo no que se refere a depreciação. Como foi visto, estas técnicas apresentam em muitas vezes erros tanto para sobrevalorização de ativos como subvalorização. Outro problema que surge é quanto a atual economia globalizada. Compras, vendas e investimentos são realizados já com grande percentual, via comércio eletrônico (e-commerce) ou negócios eletrônicos (e-business).



Essa inovadora forma de comércio e negócios exige que a situação de uma empresa seja conhecida também em *real time* (tempo real) uma vez que toda transação é realizada *on line*. Os métodos clássicos não permitem uma atualização com boa acurácia e na velocidade que o sistema eletrônico exige.

Todo empresário ou *controller* deve ter esse cuidado, uma vez que estimativas muito atualizadas indicam que num futuro não muito distante, as vendas e compras através da rede mundial *internet* superarão o comércio clássico. Estimativas apontam que até o ano de 2001 existiam cerca de 120 milhões de servidores de *internet* (*host*), conforme mostra o gráfico na figura a seguir. E esse número tem uma tendência muito forte de aumento.



Se for imaginado que cada *host* possui cerca algumas centenas ou milhares de usuários, pode-se perceber o tamanho de consumidores a espera de realização de negócios seguros e rápidos. Então, como estimar valores econômicos na velocidade exigida pelo mercado eletrônico?

4.2 O Modelo Clássico do Saldo Decrescente (Matheson)

Esse é mais um modelo clássico, mas que no entanto, permite uma relação mais iterativa (repetição) entre o valor depreciado (desconhecido) e o valor investido num ativo. O modelo assume que o valor do bem decresce a uma taxa percentual constante *p* que é avaliada no fim de cada período da vida útil. A taxa percentual é calculada pela fórmula:

$$p = 1 - \left(\frac{r}{I}\right)^{1/n}$$

onde r: retorno ou valor salvo; I: investimento; n: número de períodos de vida útil. Nestas condições, a quota de depreciação e do valor contábil no período **k+1** é obtida pelas expressões:

$$D_{k+1} = p \times I_k$$

$$I_{k+1} = I_k - D_{k+1}$$

Como se percebe, as duas variáveis (depreciação e investimento) estão intimamente e iterativamente relacionadas pelas expressões. O valor de investimento futuro depende da depreciação futura, que não é conhecida, apenas é estimada pela primeira equação. O problema fica ainda maior se lembrarmos o fato de que o fluxo de caixa sofre flutuações não previsíveis na maioria das vezes, flutuações aleatórias. Como estimar então os valores da variável depreciação se o investimento é corrompido por alterações as vezes sem explicações determinísticas?

4.3 O Valor da Informação em *Real Time* – Filtro de kalman

O filtro de Kalman é um estimador de variáveis baseado em medidas atuais (processadas em tempo real). Graças a estimativa de máxima verossimilhança com atualizações Bayesianas (probabilidades) e métodos dos mínimos quadrados (ajuste de curva), o filtro fornece uma projeção do estado "futuro" de maneira ótima. Ele é conhecido como um ótimo estimador uma vez que minimiza os erros existentes entre o valor projetado e o valor real medido.

O estimador de Kalman é também conhecido como "filtro de Kalman" (o processo com "filtragem de Kalman") uma vez que, com uma realimentação de medidas em tempo real, ele retira componentes ruidosas das medidas processadas em tempo real. Existem diversas versões para o filtro de Kalman. Nesse texto será considerado o uso do filtro de Kalman discreto. Dado um sistema dinâmico <u>linearizado</u> e <u>discretizado</u>,

$$\begin{cases} x_{k+1} = A.x_k + C.\omega_k \\ z_k = H.x_k + G.v_k \end{cases}$$

onde x é uma variável ou conjunto de variáveis que representam um problema e z uma medidas de uma variável ou um conjunto de medidas das variáveis. Com $z_0 = 0$ e x_0 dado (condição inicial do sistema) com as seguintes estatísticas dos ruídos:

$E\{\omega_k.\omega_j\} = 0$ *média do ruído de estado não correlacionado.*

$E\{\omega_k.\omega_k\} = Q$ *média do ruído de estado correlacionado.*

$E\{v_k.v_k\} = R$ *média do ruído de medida correlacionado.*

Deseja-se encontrar uma estimativa futura de uma variável x "dado" que nos são fornecidas informações passadas, representadas matematicamente pela expressão de probabilidade $\hat{x}_k = E\{x_k|z_0, z_1, ..., z_k\}$. A barra no operador de média E(|) indica "dados que". O filtro baseia-se nas equações de propagação e atualização das variáveis. Propagar significa " fazer previsão" e para tanto pode-se observar que é necessário o conhecimento de algum modelo matemático que represente o problema. A propagação será então a utilização desse modelo desconsiderando as flutuações aleatórias dos ruídos. A propagação é a utilização de um modelo clássico.

<u>Equação de Propagação:</u>

$\hat{x}_k^- = A.\hat{x}_{k-1}^+$ *propagação da variável (ou variáveis) do problema*

$P_k^- = A.P_{k-1}^+.A^T + C.Q.C^T$ *propagação da covariância do sistema*

A matriz de covariância "P" tem o mesmo sentido de representação da matriz de covariância que é estudada em estatística. Essa medida importante é conhecida como covariâncias dos dados amostrados. A matriz de covariâncias fornece uma informação compactada da correlação entre diversas variáveis envolvidas dentro e suas respecitivas variâncias. Em termos matemáticos a matriz de covariâncias pode ser apresentada da seguinte forma:

$$P = \begin{pmatrix} \sigma_1^2 & r_{12} \times \sigma_1 \times \sigma_2 \\ r_{12} \times \sigma_1 \times \sigma_2 & \sigma_2^2 \end{pmatrix}$$

onde r_{12} é o coeficiente de correlação entre as variáveis 1 e 2. Os elementos da diagonal principal são as variâncias de cada variável (por exemplo depreciação e investimento) e os elementos das diagonais secundárias fornecem a relação existente entre as duas variáveis. O coeficiente de correlação possui a clássica representação

$$r_{12} = \frac{n\sum_{i=1}^{n}(x_i y_i) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}}$$

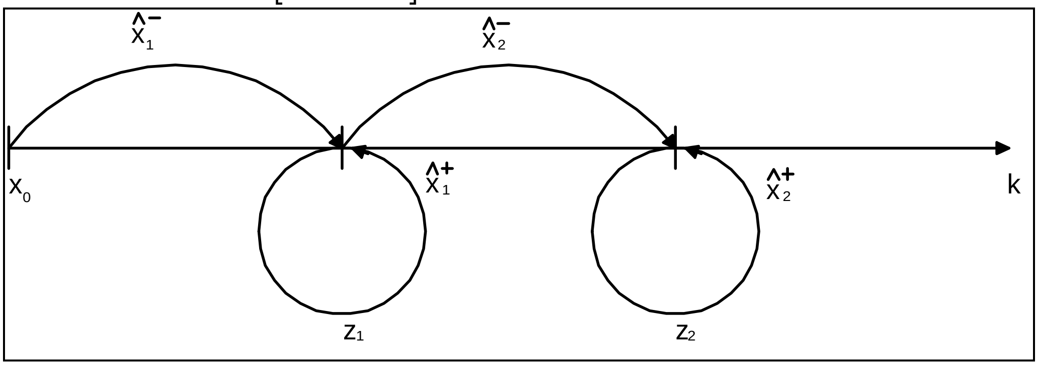
Essa é a grande vantagem do filtro de Kalman, pois de maneira iterativa e com repetições seguidas, é possível o cálculo da matriz de covariâncias P sem a necessidade de esperar pelo término de um experimento. As equações anteriores para o coeficiente de correlação não são utilizadas e já estão embutidas na recorrência das fórmulas de propagação e atualização das informações.

Equação de Atualização:

$K_k = P_k^-.H^T.\left(H.P_k^-.H^T + R\right)^{-1}$ *ganho do filtro*

$P_k^+ = P_k^- - K_k.H.P_k^-$ *atualização da matriz de covariância*

$\hat{x}_k^+ = \hat{x}_k^- + K_k.\left[z_k - H.\hat{x}_k^-\right]$ *atualização da variável(ou variáveis)*



Esquema do Filtro de Kalman para Propagações e Atualizações

4.4 Filtro de kalmam no Modelo de Saldo Decrescente

Como citado antes, a importância do filtro de Kalman é que em tempo real, pode-se estimar valores financeiros e "retirar" as flutuações de fluxo de caixa existente para previsões. Se no modelo de Saldo Decrescente apresentado antes, aparecer uma flutuação aleatória no investimento, seu poder determinístico de previsão será nulo. Vamos imaginar que essas flutuações alterem as equações do modelo para as seguintes novas relações:

$$D_{k+1} = p \times I_k$$
$$I_{k+1} = I_k - D_{k+1} + w_k$$

onde a nova variável w_k representa perturbações aleatórias nas entradas mensais, diárias, em minutos ou segundos no fluxo de caixa. O filtro de kalman discreto exige que:

- O sistema que representa o modelo seja linear.
- O sistema seja discreto.
- As perturbações aleatórias tenham médias nulas e variâncias finitas.
- As entradas nesse sistema sejam aditivas e NÃO multiplicativas.
  Assim, para o modelo proposto, w_k terá que possuir uma distribuição conhecida como Gaussiana ou Normal, com média nula e variância constante e finita do tipo,

$$w_k \sim N(0, Q)$$

Então, sem perda de generalidade, vamos adotar as variáveis depreciação e investimento como possuindo seguintes representações:

$$x_1(k) = D_k$$
$$x_2(k) = I_k$$
$$w(k) = w_k$$

O sistema então será representado por:

$$\begin{cases} x_1(k+1) = p \times x_2(k) \\ x_2(k+1) = -x_1(k) + x_2(k) + w(k) \end{cases}$$

e como a única medida possível de ser realizada em tempo real é o investimento, a variável de Kalman z_k será representada por:

$$z(k) = x_2(k) + v(k)$$

onde v(k) é a perturbação aleatória que representa o erro na hora de coletar dados sobre o investimento e terá a distribuição

$$v(k) \sim N(0, R)$$

Para ficar de acordo com a representação matricial do filtro de Kalman, o conjunto de equações anteriores pode ser visualizado da seguinte forma:

$$\begin{pmatrix} x_1(k+1) \\ x_2(k+1) \end{pmatrix} = \begin{pmatrix} 0 & p \\ -1 & 1 \end{pmatrix}\begin{pmatrix} x_1(k) \\ x_2(k) \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} w(k)$$

$$z(k) = \begin{pmatrix} 0 & 1 \end{pmatrix}\begin{pmatrix} x_1(k) \\ x_2(k) \end{pmatrix} + v(k)$$

Então, para esse modelo de depreciação, as identificações das matrizes será:

$$\begin{cases} x(k+1) = A.x(k) + C.\omega(k) \\ z(k) = H.x(k) + G.v(k) \end{cases}$$

$$x(k) = \begin{pmatrix} x_1(k) \\ x_2(k) \end{pmatrix}$$

$$A = \begin{pmatrix} 0 & p \\ -1 & 1 \end{pmatrix}$$

$$C = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$H = \begin{pmatrix} 0 & 1 \end{pmatrix}$$

$$G = 1$$

4.5 Aplicação do Filtro de Kalman

Vamos supor que uma empresa realizou uma estatística das flutuações de investimento (x_2) e percebeu que essas flutuações tinham média nula e variância Q = R\$1,00. Percebeu que seu sistema de medidas dos investimentos realizados pelos acionistas tinha uma flutuação média nula mas com variações R = R\$2,00. Quando começou a acompanhar as depreciações passadas e os investimentos passados, percebeu que a depreciação sempre tinha grandes alterações dependendo do período em que se tomava os dados históricos para observação. A variação chegava a quase 200% (P_{11} = 2) para a depreciação e para o investimento a variação era de cerca de R\$3,00 (P_{22} = 3). Não há conhecimento sobre a relação entre investimento e depreciação nessa empresa ($P_{12} = P_{21}$ = 0).

Vamos ainda considerar que no início de cada período de avaliação a depreciação inicial é nula (bastante plausível, $x_1(0) = 0$) e que o investimento inicial da empresa ou de seus acionistas é de R\$10,00 ($x_2(0) = 10$). Por último, vamos supor que o retorno (investimento salvo) é de R\$1,00 (r =1) e que o período de exercício é de 229 dias (n = 229). A empresa deseja em tempo real saber o valor da depreciação.

A resolução deste problema começa pela estimativa de “p” para o modelo de saldo decrescente. Assim,

$$p = 1 - \left(\frac{1}{10}\right)^{1/229} = 0{,}01$$

Então a formulação do problema para o uso do filtro de Kalman será:

$$\begin{pmatrix} x_1(k+1) \\ x_2(k+1) \end{pmatrix} = \begin{pmatrix} 0 & 0{,}01 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} x_1(k) \\ x_2(k) \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} w(k)$$

onde w(k) ~ N(0,1).

$$z(k) = \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} x_1(k) \\ x_2(k) \end{pmatrix} + v(k)$$

onde v(k) ~ N(0,2)

Condição inicial para depreciação e investimento

$$\begin{cases} x_1(0) = 0 \\ x_2(0) = 10 \end{cases}$$

Condição inicial para a matriz de covariâncias P

$$P(0) = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$$

Vamos imaginar, que a coleta de dados é realizada ao fim de cada dia, e que em 10 dias os resultados de investimentos foram os seguintes:

Dias	*Investimentos (R$)*
0	10
1	9,125
2	7,528
3	14,702
4	8,248
5	10,313
6	13,828
7	14,698
8	15,747
9	12,626
10	15,174

Assim, essa tabela é uma espécie de simulação e no dia 1, com investimento de R$9,125 deseja saber qual é a depreciação. No fim do dia 2 o investimento foi de R$7,528 e assim por diante. Ou seja, deseja-se saber em tempo real, para cada valor amostrado de investimento qual a depreciação.

Solução

Para o primeiro dia após o investimento inicial, a medida foi de R$9,125. Então para descobrir a depreciação, faz-se primeiro k = 0 (iteração). Então

k =0
<u>Propagação</u>

$$P(1)^- = \begin{pmatrix} 0{,}0003 & 0{,}03 \\ 0{,}03 & 6 \end{pmatrix}$$

$$\hat{x}(1)^- = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

<u>Atualização</u>

$$K(1) = \begin{pmatrix} 0{,}00375 \\ 0{,}75 \end{pmatrix}$$

$$P(1)^+ = \begin{pmatrix} 0{,}00019 & 0{,}0075 \\ 0{,}0075 & 1{,}5 \end{pmatrix}$$

$$\hat{x}(1)^+ = \begin{pmatrix} 0{,}0375 \\ 7{,}5 \end{pmatrix}$$

---

k =1
<u>Propagação</u>

$$P(2)^- = \begin{pmatrix} 0{,}00015 & 0{,}0149 \\ 0{,}0149 & 2{,}485 \end{pmatrix}$$

$$\hat{x}(2)^- = \begin{pmatrix} 0{,}075 \\ 7{,}462 \end{pmatrix}$$

<u>Atualização</u>

$$K(2) = \begin{pmatrix} 0{,}0033 \\ 0{,}554 \end{pmatrix}$$

$$P(2)^+ = \begin{pmatrix} 0{,}00010 & 0{,}0066 \\ 0{,}0066 & 1{,}108 \end{pmatrix}$$

$$\hat{x}(2)^+ = \begin{pmatrix} 0{,}0805 \\ 8{,}384 \end{pmatrix}$$

---

k =2
<u>Propagação</u>

$$P(3)^- = \begin{pmatrix} 0{,}00011 & 0{,}01102 \\ 0{,}01102 & 2{,}094 \end{pmatrix}$$

$$\hat{x}(3)^- = \begin{pmatrix} 0{,}0838 \\ 8{,}303 \end{pmatrix}$$

Atualização

$$K(3) = \begin{pmatrix} 0{,}00269 \\ 0{,}5116 \end{pmatrix}$$

$$P(3)^{+} = \begin{pmatrix} 0{,}00008 & 0{,}00538 \\ 0{,}00538 & 1{,}023 \end{pmatrix}$$

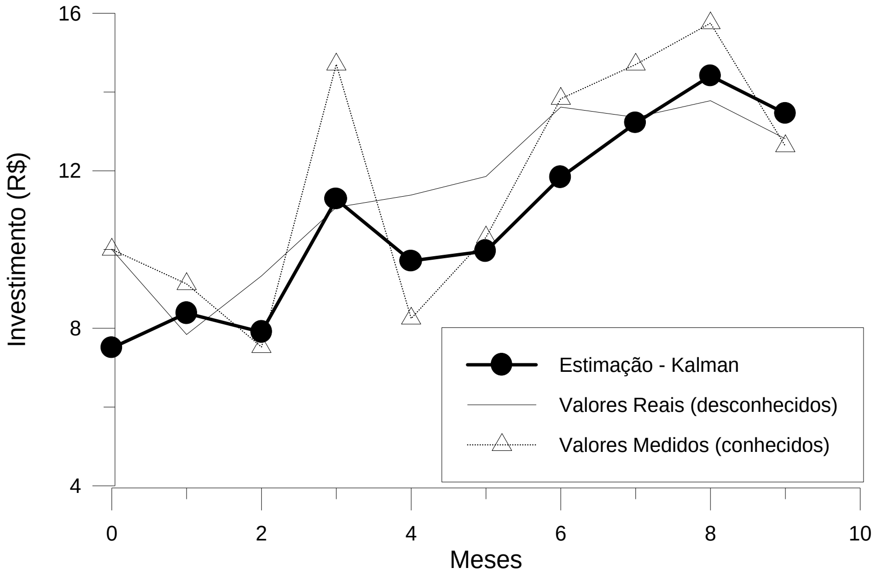
$$\hat{x}(3)^{+} = \begin{pmatrix} 0{,}0817 \\ 7{,}906 \end{pmatrix}$$

O leitor poderá perceber que no primeiro dia após o investimento inicial, foi previsto uma entrada de investimento de R$7,5 para o dia seguinte. Se observar a tabela “real ” verá que no seguinte a entrada foi de R$7,528 ! Um erro de R$0,028 ! A depreciação então pelo filtro de Kalman forneceu 3,75% (0,0375). Esses valores então entram para realimentar as equações no passo k = 1 (para encontrar $\hat{x}(2)^{+}$) e assim sucessivamente.

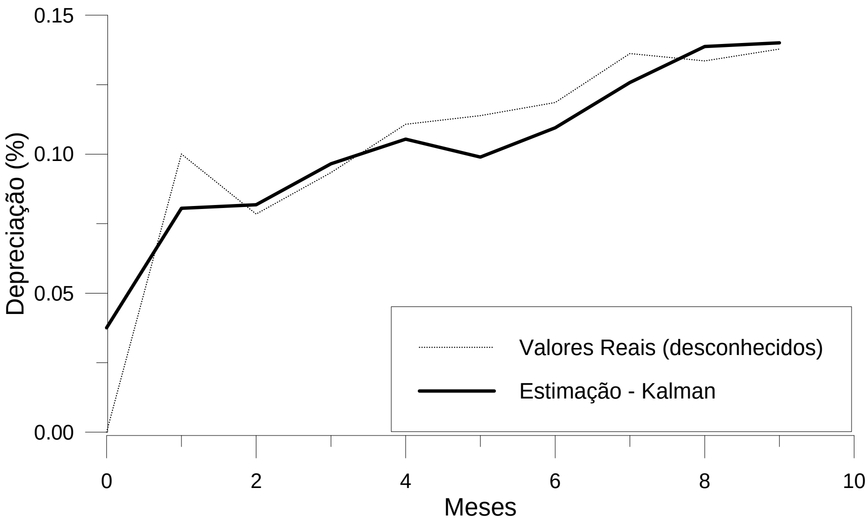
Os gráficos a seguir mostram a comparação entre os valores reais medidos, valores reais desconhecidos (mas estimados pelo modelo matemático do saldo decrescente) e os valores estimados pelo filtro de Kalman.

Pode-se perceber então a grande utilidade do filtro de Kalman, uma vez que não só filtra dados reais conhecidos (investimento, por exemplo) como também estima dados não possíveis de medir em tempo real (depreciação, por exemplo).

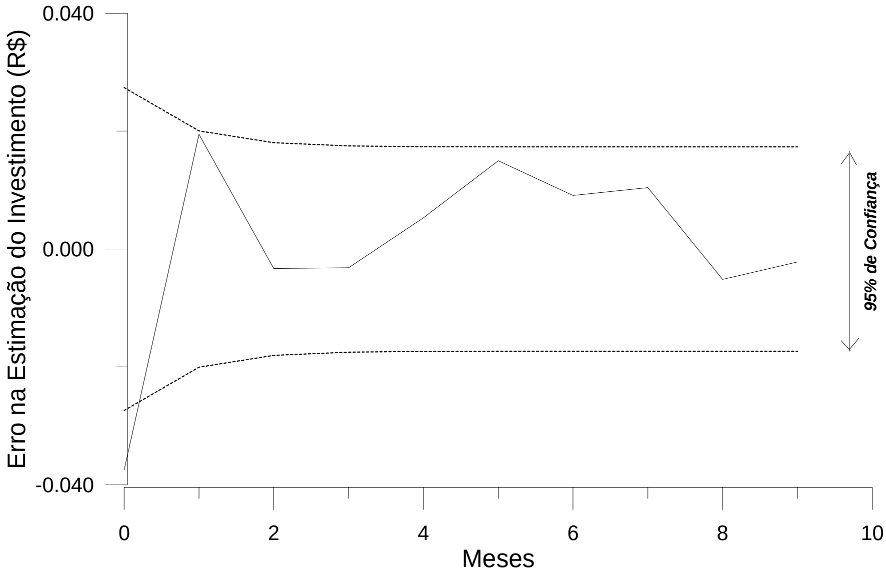
Uma palavra a mais de comentário sobre a matriz de covariância deve ser escrita. Essa matriz P (matriz de covariância) mede a memória do filtro de Kalman. A medida que valores novos vão entrando para os cálculos futuros, o filtro “vai perdendo” memória em relação aos efeitos dos dados passados. Toda vez que uma exponencial decrescente representa a matriz P, é um bom sinal de ajuste do filtro, pois ele dá mais valor a informação nova. Caso contrário, o filtro não estará funcionando de acordo, ou porque o modelo não é adequado ao problema, ou as medidas possuem uma variação maior do que a estimada, ou ainda pela instabilidade do filtro em relação aos cálculos.

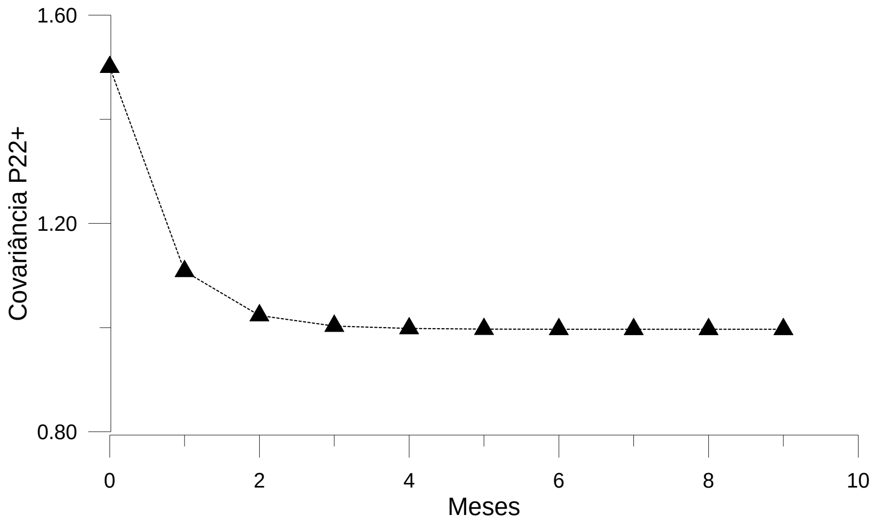


16
12
8
4
Investimento (R$)
0
2
4
6
8
10
Meses
Estimação - Kalman
Valores Reais (desconhecidos)
Valores Medidos (conhecidos)



0.15
0.10
0.05
0.00
Depreciação (%)
0
2
4
6
8
10
Meses
Valores Reais (desconhecidos)
Estimação - Kalman





4.6 Conclusão

Os modelos de depreciação sempre são bastante discutíveis e não existe um modelo perfeito para ampla utilização. Os modelos clássicos, como os apresentados, são bastante limitados como já comentados antes. Em se tratando de comércio eletrônico tornam-se ainda mais obsoletos. Conforme a revisão de estatística mostrou, é de fundamental importância a coleta confiável de dados. A confiabilidade fornecerá um ajuste bastante preciso em termos de adequação dos parâmetros aos modelos escolhidos.

O filtro de kalman surge como mais uma alternativa para a estimativas rápidas e confiáveis sobre valores econômicos. Concebido na década de 1960 para a estimação de trajetórias espaciais, tem-se mostrado útil nesses 40 anos em diversas áreas, tais como: medicina, economia, aeronáutica, oceanografia, matemática, etc. Em finanças existem diversos trabalhos internacionais em anos mais recentes utilizando filtro de Kalman. Este fato atesta, que em mais essa área essa ferramenta torna-se poderosa para comércios e negócios eletrônicos via *internet*.