ANÁLISE DE RISCO EM APLICAÇÕES FINANCEIRAS

MARCO ANTONIO LEONEL CAETANO

Blucher

Marco Antonio Leonel Caetano

ANÁLISE DE RISCO EM APLICAÇÕES FINANCEIRAS

Análise de risco em aplicações financeiras

Editora Edgard Blücher Ltda.

Blucher

Rua Pedroso Alvarenga, 1245, 4° andar
04531-934 – São Paulo – SP – Brasil
Tel.: 55 11 3078-5366
contato@blucher.com.br
www.blucher.com.br

Segundo Novo Acordo Ortográfico, conforme 5. ed. do *Vocabulário Ortográfico da Língua Portuguesa*, Academia Brasileira de Letras, março de 2009.

FICHA CATALOGRÁFICA

Caetano, Marco Antonio Leonel
Análise de risco em aplicações financeiras / Marco Antonio Leonel Caetano. — São Paulo: Blucher, 2017.
264 p. ; il.

Bibliografia
ISBN 978-85-212-1144-0

1. Finanças – Investimentos – Avaliação de riscos 2. Administração de risco - Matemática 3. Sistemas de avaliação de risco de crédito (Finanças) 4. Mercado financeiro 5. Probabilidades I. Título.

15-1526 CDD 332.604

Índice para catálogo sistemático:
1. Investimentos – Avaliação de riscos

CONTEÚDO

CAPÍTULO 1
TRATAMENTO, QUANTIFICAÇÃO E VISUALIZAÇÃO DE DADOS

1.1 INTRODUÇÃO

Na divulgação de relatórios, estrategistas e analistas preocupam-se com a melhor maneira de apresentar os dados obtidos por meio dos resultados das operações nos investimentos financeiros. Necessariamente, toda divulgação deve começar sempre pela divulgação dos resultados com técnicas estatísticas adequadas.

Na estatística, existe um amplo campo de técnicas para expressar e representar resultados de investimentos em três subáreas bastante distintas e muito bem conectadas: estatística descritiva, probabilidade e inferência estatística. Na estatística descritiva são utilizadas ferramentas como objeto de apresentação, como gráficos, tabelas e análises de formas e estruturas de representações dos dados. Esse tipo de estatística não tem valor de inferência, ou seja, as conclusões são sempre limitadas, e também não permite a extensão direta dos resultados para a população da qual foram coletadas as amostras. No campo da probabilidade, teoremas garantem os tipos de distribuição que regem um evento ocorrido no mercado, destacam quais são as distribuições mais adequadas e mostram com que confiança os dados podem ser coletados de forma que sejam representativos de uma população de dados. Finalmente, na inferência estatística, técnicas garantem as conclusões indicando os erros e as confiabilidades dos resultados obtidos. Também nesse campo é possível fazer previsões de tendências e correlações entre variáveis e parâmetros obtidos experimentalmente.

Desse modo, realizar uma projeção ou divulgação de relatórios de investimentos sem o devido cuidado com seu tratamento e forma de apresentação pode comprometer todo trabalho realizado, por falta de compreensão por parte dos leitores ou de interpretações incorretas de determinados resultados obtidos.

1.2 REPRESENTAÇÃO GRÁFICA

A melhor forma de representar dados é com gráficos, que muitas vezes, por si sós, são bastante explicativos e conclusivos dependendo do evento observado. Para exemplificar alguns tipos de gráficos, são usados os dados fornecidos a seguir das exportações brasileiras no ano de 2001 em milhões de dólares. Esses dados foram obtidos mês a mês.

Tabela 1.1 – Exportações (total em milhões de dólares – US$)

Mês	1	2	3	4	5	6	7	8	9	10	11	12
US$	4.538	4.083	5.167	4.730	5.367	5.042	4.965	5.727	4.755	5.003	4.500	4.346

Fonte: Ministério do Desenvolvimento, Indústria e Comércio Exterior (2001).

1.2.1 GRÁFICO DE LINHAS

É uma das formas mais simples de apresentação. Esse tipo de 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 as tendências dos resultados obtidos. A Figura 1.1 apresenta um gráfico típico de linhas.

Figura 1.1 – Gráfico de linhas com dados das exportações brasileiras em 2001.

1.2.2 GRÁFICO DE PONTOS

Forma de gráfico bastante interessante quando há frequências nos dados coletados. Nesse caso, o leitor consegue visualizar o valor que mais se repete em uma amostragem. Os dados do exemplo são referentes à frequência da exportação brasileira entre 1999 e 2001 observada nos dias úteis do mês.

Tabela 1.2 – Frequência de dias úteis do mês (1999-2001)

Dias úteis	Frequência
18	3
19	2
20	12
21	10
22	10
23	4

Fonte: Ministério do Desenvolvimento, Indústria e Comércio Exterior (2001).

Para elaborar o gráfico, são colocados pontos nos valores observados, seguindo no eixo horizontal e na frequência com que aparecem na amostra no eixo vertical. A Figura 1.2 apresenta o gráfico de pontos.

Figura 1.2 – Gráfico de pontos mostrando dias úteis.

1.2.3 HISTOGRAMA

Esse tipo de gráfico consiste em retângulos justapostos que indicam, em sua base, o intervalo dos valores de dados do experimento cuja frequência é representada pela altura do retângulo. O sentido é um pouco mais amplo que o do gráfico de pontos, pois o interesse não está em um único valor, mas sim em um intervalo de valores amostrados.

Os valores a seguir correspondem à média diária mensal de importações brasileiras de janeiro de 1999 a janeiro de 2002.

Tabela 1.3 – Média diária mensal de importações (em milhões de US$)

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 (2001).

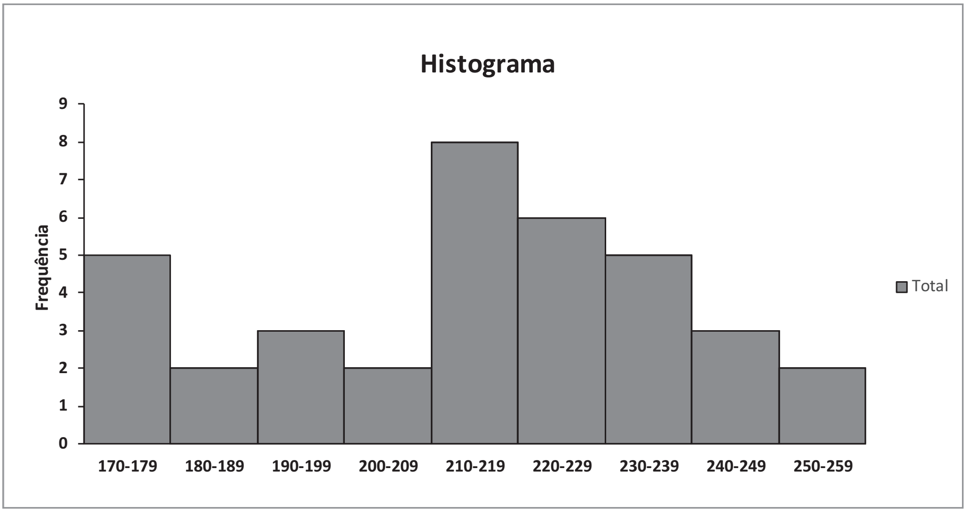


Figura 1.3 – Histograma dos valores de importações brasileiras (1999-2002).

A figura apresenta o histograma com os valores da Tabela 1.3. Pode-se notar alguns fatos interessantes, por exemplo, auxilia a interpretar que o valor mais frequente de importações está entre US$ 210 milhões e US$ 219 milhões em termos de média diária.

1.2.4 GRÁFICO DE BARRAS

Um gráfico de barras, assim como o histograma, representa os valores obtidos no experimento em termos de frequência para cada valor observado. A diferença é que esse gráfico não é utilizado para intervalos amostrados, mas sim para os valores observados de maneira individual. A Figura 1.4 representa os valores das importações brasileiras da Tabela 1.3.

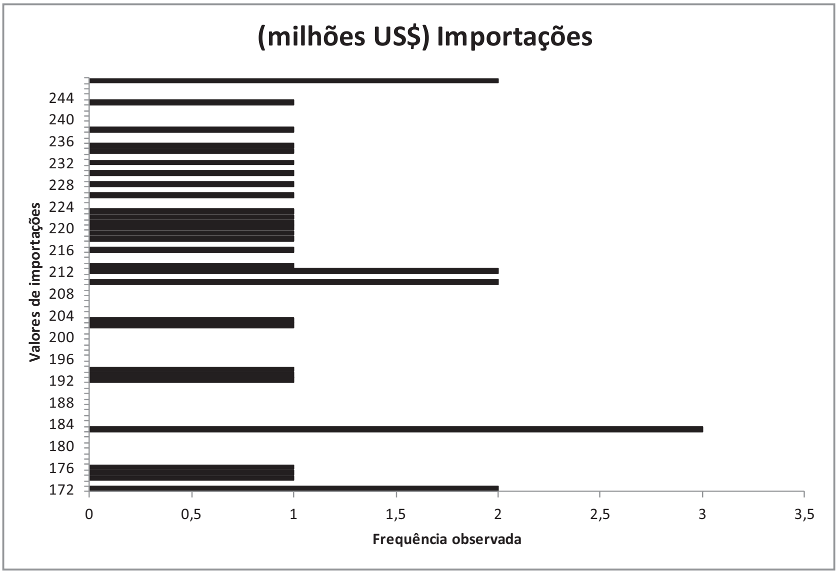


Figura 1.4 – Gráfico de barras das importações brasileiras.

1.2.5 CURVAS DE NÍVEL

As curvas de nível são um tipo bastante interessante de gráfico, pois apresentam isolinhas para os pontos amostrados. Isso significa que, uma vez que o valor é escolhido em uma das linhas, todos os pontos para as posições x e y são iguais para uma variável z que seja função de x e y. Em uma curva em três dimensões estimada por uma função, a representação é $z = f(x,y)$.

EXEMPLO 1.1

Tabela 1.4 – Orçamento federal (milhões de reais – R$)

Administração	Saúde
6.532	13.219
18.091	11.836
19.442	10.986
19.740	10.220

Fonte: Sistema Integrado de Administração Financeira – Coordenação Geral de Contabilidade e Custos da União/Secretaria do Tesouro Nacional (2003).

Essa tabela apresenta os valores da execução orçamentária das despesas federais com administração e saúde de janeiro a setembro de 1996 até 1999. Supõe-se que uma curva de ajuste boa para a relação entre gastos com administração e gastos com saúde seja:

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

em que x é o gasto com administração e y, o gasto com saúde.

Nesse caso, a variável z expressa uma relação não linear entre os dois tipos de gastos do governo. Uma aplicação pode advir de uma medida estatística sobre a correlação não linear dos dados. Essa representação ajudaria a entender o que acontece com a relação entre as duas variáveis (x,y) quando um dado em x aumenta e outro em y diminui. Os dados reais são apresentados com o orçamento em milhões de reais obtidos do *site* do Instituto de Pesquisa Econômica Aplicada (Ipea), a partir de dados mensais da fonte do governo federal, como mostra a tabela. As isolinhas para esses valores são as formas traçadas no gráfico da Figura 1.5.

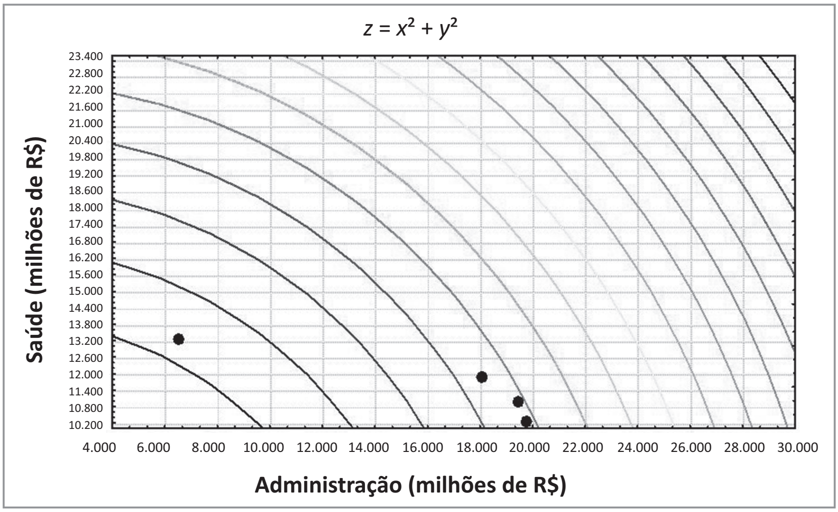


Figura 1.5 – Isolinhas dos gastos federais representando a forma da função $z = f(x,y)$.

Essa representação gráfica é bastante útil no oitavo capítulo, quando o risco de investimentos no mercado financeiro é relacionado com um índice baseado na técnica wavelet e com a volatilidade nos preços dos ativos. As isolinhas no estudo do risco estão relacionadas com a probabilidade de quedas mais fortes no preço dos ativos.

1.3 MEDIDAS DESCRITIVAS DOS DADOS

Na seção anterior, foram apesentadas formas gráficas de representação de dados de um relatório empresarial. Cabe ao gestor escolher e adequar a melhor forma de apresentação de seus resultados, de forma que elucide todos os fatos com uma simples visualização dos acontecimentos. No entanto, na maioria das vezes, essa facilidade não é obtida 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. Cabe, então, fazer uso de formas quantitativas de extração de informações por meio de medidas estatísticas que apresentem, de maneira rápida e sucinta, as inter-relações existentes no fenômeno em estudo. O gestor deve escolher como as variáveis vão representar os principais fatores decorrentes do experimento.

Uma variável pode ser discreta ou contínua, dependendo do tipo de estudo executado. Uma variável discreta X é aquela em que o número de valores possíveis de X é finito ou infinito enumerável. Ou seja, os valores possíveis de X podem ser colocados em uma fila do tipo x_1, x_2, x_3 ... x_n. Geralmente, esse tipo de variável é utilizado em problemas de contagem (COSTA NETO, 1999). São exemplos as contagens de firmas em concordata, nível de emprego, contagem do número de vagas abertas por uma empresa etc. Para uma variável contínua Y, como o próprio nome diz, os dados podem até ser observados de forma discreta, mas os valores de Y pertencem ao conjunto dos números reais.

EXEMPLO 1.2

A Tabela 1.5 apresenta um exemplo de variáveis contínuas.

Tabela 1.5 – Taxa de desemprego no Brasil (janeiro de 1999 a maio de 2001, em %)

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: Fundação Seade. ■

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

O primeiro tratamento representativo para extração de informação dessa coleta é por meio 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 e são bastante comuns na literatura nacional (COSTA NETO, 1999; FONSECA, 1992; SPIEGEL, 1994).

- Dados brutos (n): dados ainda não organizados, como na Tabela 1.5.
- Rol: é o arranjo dos dados brutos em ordem crescente ou decrescente.
- *Range* ou amplitude total: é a diferença entre o maior e o menor valor observado.
- Frequência absoluta da classe (F): número de vezes que o elemento aparece na amostra ou o número de elementos pertencentes a uma classe.
- Número de classes (k): existem algumas maneiras de determinar um número adequado de classes, entre elas estão estas duas:

  (a) número será $k = 5$, se o número de dados for menor ou igual a 25.

  (b) para número de dados superior, utiliza-se $k = \sqrt{n}$.
- Amplitude das classes (h): a amplitude pode ter larguras diferentes, mas, para efeitos práticos, podem ser adotadas como de mesmo tamanho utilizando-se esta definição:

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

- Limite das classes: limite inferior (Li); limite superior (Ls).

  Li |-----| Ls: compreende os valores Li e Ls;

  Li |----- Ls: não compreende o valor Ls;

  Li -----| Ls: não compreende o valor Li.
- Ponto médio das classes (PM_i): média dos valores limitantes das classes:

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

- Frequência absoluta acumulada direta (Fac): é a soma das frequências absolutas de valores inferiores ou iguais ao valor da frequência da classe.
- Frequência 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, como na Tabela 1.6 (tabela de classes), o nível de desemprego no país.

Tabela 1.6 – Tabela de classes para nível de desemprego no Brasil (janeiro de 1999 a maio de 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

Classes	*F*	*f*	*PM*	Fac
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: Fundação Seade.

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

1.3.1 MEDIDAS DE POSIÇÃO

São definidas de modo que apresentam 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 média, mediana e moda.

Média aritmética

- 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 é:

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

- Dados agrupados em tabela de frequência: sejam $x_1, x_2, ..., x_n$ pontos médios das classes construídas para os dados brutos. Sendo as frequências $F_1, F_2, ..., F_n$ de cada classe, a média para os dados agrupados é calculada por:

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

EXEMPLO 1.3

Tabela 1.7 – Cheques sem fundos (média/1.000)

Devolução	Frequência 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 tabela representa o número de cheques sem fundos devolvidos uma segunda vez em cada mil cheques apresentados, de maio de 2001 a maio de 2002. A média aritmética para essa tabela é de 14,1 cheques entre maio de 2001 e maio de 2002 para cada mil apresentados.

- Dados agrupados em tabela de classes: as classes são representadas pelos seus pontos médios, conforme a Tabela 1.6. Nesse caso, a média é calculada por:

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

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

Mediana

Um valor é dito mediano quando divide o conjunto de dados do experimento depois de sua ordenação, separando-o em dois subconjuntos com igual número de elementos. Sua notação em geral é $\widetilde{x}$:

- Dados não agrupados:
  5 7 8 10 14 A mediana é $\widetilde{x} = 8$
  5 7 8 10 14 15 A mediana é $\widetilde{x} = 9$

Uma maneira de encontrar a mediana de um conjunto composto de dados brutos seria esta: se o número de dados n é ímpar, a mediana é o elemento central $(n + 1)/2$ da amostra ordenada, caso contrário, a mediana é a média dos elementos centrais dessa amostra ordenada formada por $[n/2, (n/2) + 1]$.

- Dados agrupados por frequência: nesse caso, cria-se uma nova coluna de frequências acumuladas diretas para auxílio na escolha da mediana. A Tabela 1.7 se transforma na Tabela 1.8, a seguir.

Tabela 1.8 – Cheques sem fundos (média/1.000)

Devolução	Frequência absoluta	Frequência acumulada
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.

Nesse caso, $n = 13$ é ímpar. Então, $(n + 1)/2 = 7$, o que significa que o sétimo elemento corresponde ao elemento mediano desse conjunto de valores de cheques devolvidos. Logo, a mediana é 13,6, diferente da média, que é 14,1.

- Dados agrupados em tabela de classes: nesse caso, é necessária uma fórmula de interpolação para se encontrar o elemento mediano. Deve-se ressaltar que esse valor é apenas representativo e que não necessariamente faz parte da amostra. Os passos a seguir são:
  (a) Calcula-se a ordem $(n/2)$, não se preocupando se for par ou ímpar, pois a variável é contínua. A classe da mediana é aquela cuja frequência acumulada até ela é maior ou igual a $(n/2)$, já a imediatamente anterior é menor que $n/2$.
  (b) Utiliza-se a seguinte fórmula de interpolação:

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

em que L_{md} é limite inferior da classe da mediana; n representa o tamanho da amostra; $\sum f$ significa frequência acumulada da classe imediatamente anterior à da mediana;

h é amplitude da classe da mediana; e F_{md} representa a frequência 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 é:

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

Existem ainda medidas alternativas para dividir os dados em quatro, dez e cem partes iguais. 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 frequências acumuladas diretas também são trocados pelos limites das classes dos quartis, decis e percentis.

Moda

Quando a variável é discreta, a medida de moda representa o elemento mais frequente na amostragem, ou seja, aquele que mais se repete. No entanto, para o caso em que a variável é contínua, essa definição deixa de valer e passa-se a adotar uma medida de densidade.

- Dados não agrupados.

EXEMPLO 1.4

Sejam os dados de uma amostragem formada por 2, 7, 9, 5, 6, 3, 7, 4, 1, 7, a moda é o número 7. ■

- Dados em tabela de classes: da mesma forma que na mediana se faz necessária a interpolação dos dados para encontrar a moda, existem alguns tipos de fórmulas que se diferenciam no cálculo da moda para a tabela de classes. Uma dessas fórmulas pode ser usada adotando-se os seguintes passos:
  (a) identifica-se a classe modal, ou seja, aquela com a maior frequência absoluta.
  (b) utiliza-se a fórmula:

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

em que L_i é o limite inferior da classe modal; F_i representa a frequência absoluta da classe modal; F_{i-1} é a frequência absoluta da classe imediatamente anterior à classe modal; F_{i+1} significa a frequência absoluta da classe imediatamente posterior à classe modal; e h é a amplitude da classe modal.

Observando novamente a Tabela 1.6, pode-se verificar que a classe modal é a última classe com treze 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 conhecidas as medidas de posição de uma curva representativa dos dados de uma avaliação financeira ou empresarial, faz-se necessário saber se os valores numéricos dos dados coletados são mais ou menos dispersos em torno das medidas de posição. Torna-se indispensável o conhecimento da dispersão desses dados em relação às medidas de posição, principalmente em relação à média. Essa dispersão é fundamental para determinar o grau de confiança nas análises de inferência do problema e suas projeções para o quadro financeiro. São quatro as medidas a serem apresentadas.

Amplitude total

Essa medida é muito simples e constitui a 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 depende apenas da dispersão dos valores extremos, não sendo afetada pela dispersão dos valores internos.

Variância

Mede a dispersão dos dados em torno da média. A título de exemplo, suponha-se que se tenha o seguinte conjunto de dados A = {3, 4, 5, 6, 7}, em que a média é 5. Calculando-se o desvio das unidades do conjunto A em relação à média, tem-se:

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

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

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

Porém, como está, essa medida cresceria indefinidamente à 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:

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

Essa forma de medida passa a ser chamada variância populacional, uma vez que foi ponderada por todos os termos amostrados. Às 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 ser tomada é dividir as somas dos desvios não pelo total *n* de dados, mas sim por *n–1* dados. A teoria de probabilidade prova que esse é um bom "truque" de correções de tendenciosidade na amostragem. Logo, a segunda medida de variância é:

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

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

(a) variância populacional

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

(b) variância amostral

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

A variância amostral para o conjunto A descrito anteriormente é 2,5. No caso de haver dados já apresentados em tabela de frequências, o cálculo da variância pode ser realizado diretamente com:

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

em que a variável F_i representa a frequência absoluta dos dados.

EXEMPLO 1.5

A Tabela 1.7 apresenta a devolução de cheques em tabela de frequência. Para este exemplo, a média encontrada foi 14,1 e a variância amostral pode ser calculada da seguinte forma:

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

■

A última forma de apresentação da variância é quando se tem os dados em forma de tabela de classes. Nesse caso, o cálculo da variância é:

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

em que PM_i é o ponto médio de cada classe.

EXEMPLO 1.6

Utilizando-se da Tabela 1.6, foi encontrada na seção anterior a média para a tabela de classes de 7,19% de nível de desemprego. A variância amostral para esse exemplo é 0,618. ■

Desvio-padrão

Essa medida fornece ao pesquisador uma maneira de saber matematicamente a oscilação em torno dos dados. O desvio-padrão, assim como a variância, apresenta o grau de confiabilidade dos dados em torno da média e corrige um problema da variância quanto ao elevado grau de medida. Enquanto os valores da variância podem tornar-se bastante elevados, a ponderação da variância com a extração da raiz quadrada torna a medida mais tratável do ponto de vista numérico.

No mercado financeiro, é comum usar o jargão "volatilidade" para referir-se ao desvio-padrão. Sabe-se da extensa literatura sobre teoria da probabilidade (MEYER, 1984; PAPOULIS, 1991; MAGALHÃES; LIMA, 2000) que em 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

É uma medida relativa da dispersão – em porcentagem – de quanto a variabilidade influencia na confiança da média calculada. Com um coeficiente de alto grau – por exemplo, acima de 50% –, pode-se dizer que existe alto grau de dispersão dos dados em relação à média encontrada. Assim, a maneira de calcular o coeficiente de variação é:

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

EXEMPLO 1.7

Com a Tabela 1.5, pode-se saber qual o comportamento dos dados em relação à média encontrada no nível de desemprego. A média foi de 7,19% de desemprego. Sendo o desvio-padrão ±0,786, o coeficiente de variação é:

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

Sabendo que a variação é de 10,9%, é possível concluir que os dados sobre o desemprego oscilam pouco em relação à média. O *cv* indica que existe uma variabilidade de 11% com relação ao valor encontrado pela média. ■

1.3.3 MEDIDAS DE ASSIMETRIA

Esse tipo de medida é bastante útil quando se deseja saber a forma da curva que os dados da amostra apresentam. Essa curva pode ser simétrica, quando a área em relação às medidas de posição são iguais, com valores acima ou abaixo de média, mediana e moda. Se a curva é simétrica, há o caso em que a média é igual à moda e ambas são iguais à mediana. No entanto, o contrário não é garantia para a definição de igualdade. Isso significa que, muitas vezes, pode-se ter as três medidas iguais e, mesmo assim, uma curva assimétrica.

A assimetria vai ser positiva quando o coeficiente de assimetria (AS) também for positivo, indicando que o valor modal é inferior ao valor médio. A assimetria vai ser negativa quando o valor modal for maior que o valor médio. O coeficiente de assimetria pode ser calculado como:

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

em que *x* é o valor médio, *Mo* representa o valor modal e *s* é o desvio-padrão. No entanto, essa fórmula pode apresentar um inconveniente. Muitas vezes não se tem um valor modal ou há muitos valores modais. Nesses casos, uma fórmula alternativa é o coeficiente de Pearson, que faz uso do valor mediano e dos quartis na forma (FONSECA, 1992):

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

em que os *Q* representam os quartis terceiro e primeiro e o valor mediano.

1.4 MEDIDAS DESCRITIVAS DE POSIÇÃO NO EXCEL

Atualmente, as fórmulas das seções anteriores estão disponíveis e facilmente acessadas em diversos programas de computador. O mais simples, por exemplo, é utilizar as medidas para descrever diversos eventos com planilhas do Microsoft Excel.

A estatística descritiva apresenta uma foto de momento – assim como a inferência – e, por isso, é preciso ter cuidado com tentativas de extrapolação de resultados baseadas apenas em medidas descritivas ou em medidas de dispersão. Em alguns casos, essas medidas são suficientes para a compreensão dos eventos, mas é preciso ter cuidado com seu poder de previsão. Deve-se lembrar que elas apenas descrevem o fenômeno na data de observação.

Por exemplo, a Figura 1.6 apresenta os dados do Ibovespa (índice da bolsa de valores Bovespa) do ano de 2007, portanto, pré-crise financeira de 2008.

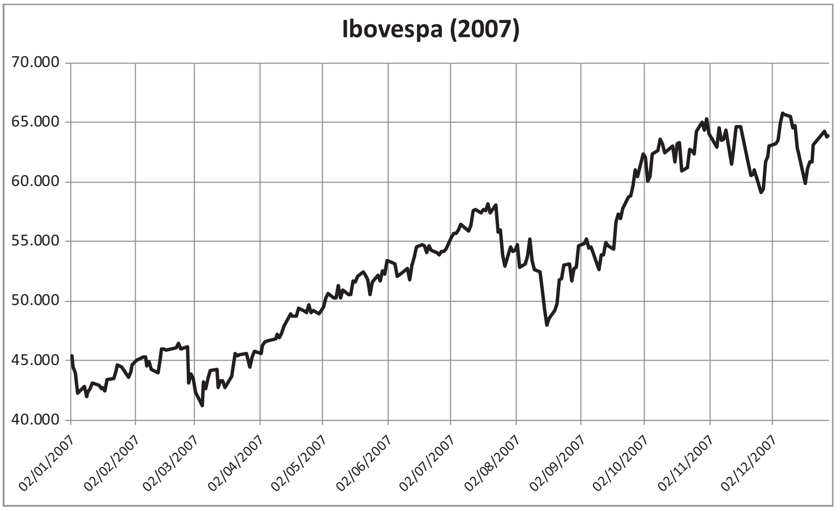


Figura 1.6 – Ibovespa em 2007.

Na figura, há 245 dados diários de fechamento do Ibovespa (valores após o fechamento do pregão na Bovespa). Os cálculos de algumas medidas descritivas são bastante fáceis ao utilizar, por exemplo, funções do Microsoft Excel. Existem algumas variações entre uma versão e outra, mas de modo geral as funções mudam pouco.

Para os dados em questão, pode-se calcular a amplitude de variação do Ibovespa utilizando as funções **Máximo( )** e **Mínimo( )**. Colocando as datas na coluna A de uma planilha do Excel e o Ibovespa na coluna B, a função **Máximo( )** fica desta forma:

E1 *fx* =MÁXIMO(B1:B245)

	A	B	C	D	E	F
1	02/01/2007	45383		Máximo	65791	
2	03/01/2007	44445				
3	04/01/2007	44020				
4	05/01/2007	42245				
5	08/01/2007	42830				
6	09/01/2007	42007				
7	10/01/2007	42336				

Figura 1.7 – Função máximo no Excel.

Do mesmo modo, a função **Mínimo( )** é representada assim:

fx =MÍNIMO(B1:B245)

O cálculo da amplitude vai ser a diferença entre o resultado do máximo e o mínimo valor do Ibovespa. No caso em questão, nossa amplitude é:

E5 *fx* =E1-E3

	A	B	C	D	E
1	02/01/2007	45383		**Máximo**	65791
2	03/01/2007	44445			
3	04/01/2007	44020		**Mínimo**	41179
4	05/01/2007	42245			
5	08/01/2007	42830		**Amplitude**	24612

Figura 1.8 – Cálculo da amplitude do Ibovespa (2007) no Excel.

As funções média e mediana são encontradas no Excel da seguinte maneira:

fx =MÉDIA(B1:B245)

fx =MED(B1:B2)

A medida de moda também está programada no Excel, por exemplo, na função **Modo.Único( )**. No caso de dados não repetitivos, porém, não tem muita utilidade. Dados como os do Ibovespa, que dificilmente se repetem por conta da composição de diversas ações na Bovespa, fornecem resultado vazio, como mostrado a seguir:

O resultado é 

.

Existe uma maneira bem mais interessante de completar as medidas descritivas para fazer estatística de dados tabelados em planilha do Excel. Uma utilização interessante é o do conceito de tabela dinâmica. Essa função existe em todas as últimas versões do Excel – na versão 2010 do Microsoft Excel está na aba Inserir, conforme Figura 1.9. O símbolo sempre muda de uma versão para outra. Aqui é apresentado o da versão 2010 (primeiro ícone à esquerda):

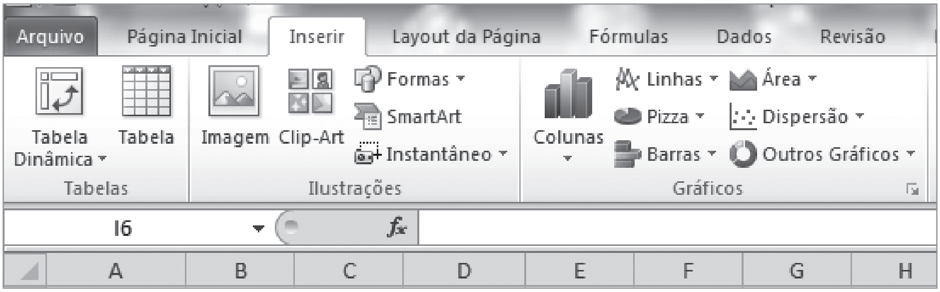


Figura 1.9 – Inserção da tabela dinâmica.

Ao clicar no símbolo da tabela dinâmica, aparece uma tela interativa solicitando o intervalo de entrada dos dados, conforme a figura a seguir.

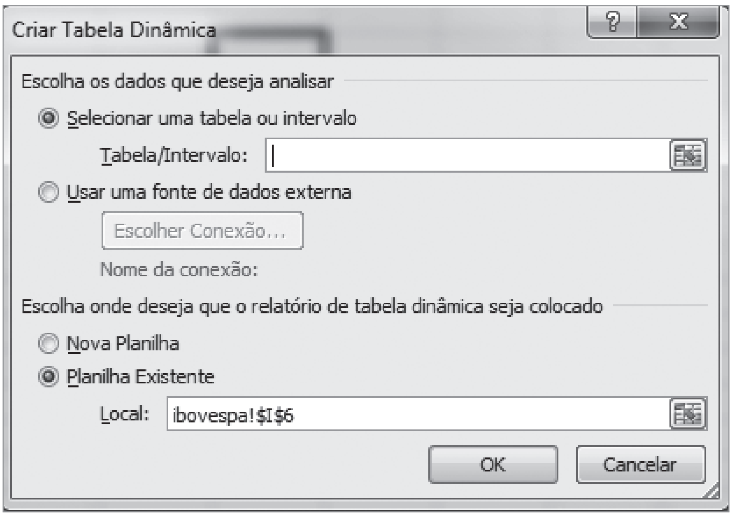


Figura 1.10 – Tela inicial da tabela dinâmica.

Escolhe-se, então, todo o intervalo em que estão os dados do Ibovespa, por exemplo, e clica-se no botão **OK**. Se os dados estiverem na coluna B, como no caso do exemplo, o preenchimento fica desta forma:

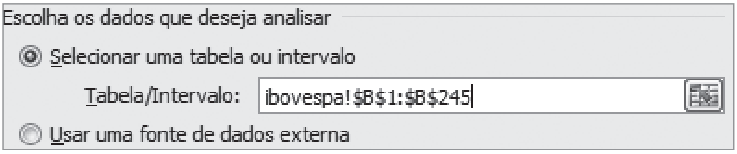


Figura 1.11 – Preenchimento do intervalo na tabela dinâmica.

O resultado que aparece é uma pequena tabela vazia e uma caixa de informação esperando detalhes do que se deseja fazer com os dados selecionados.

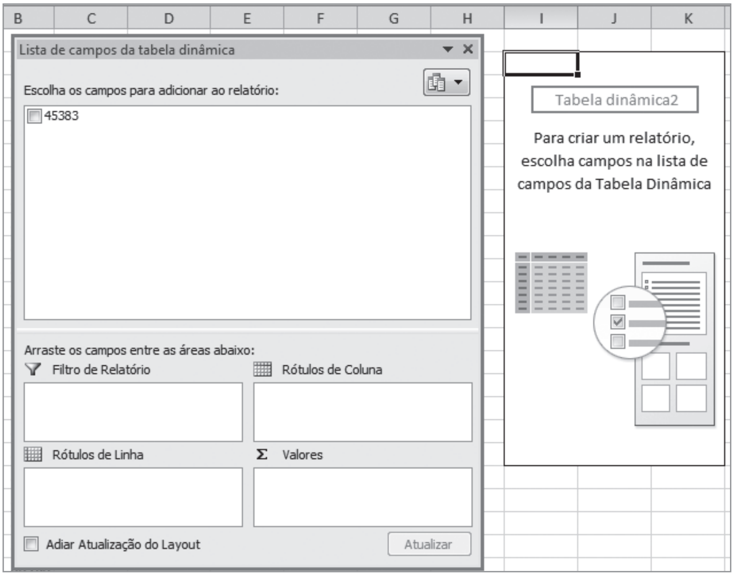


Figura 1.12 – Tabela dinâmica.

O primeiro número que aparece dentro da caixa de informação da tabela dinâmica é o primeiro valor do Ibovespa para o dia 2 de janeiro de 2007. Arrastando esse número para dentro da caixa **Rótulos de Linha** e depois para dentro da caixa **Valores**, são obtidas duas colunas iguais.

Rótulos de Linha	Soma de 45383
41179	41179
42007	42007
42245	42245
42336	42336
42370	42370
42478	42478

Figura 1.13 – Tabela dinâmica: Rótulos de Linha.

Na coluna da esquerda, é possível, com o botão da direita do *mouse*, escolher **Agrupar**. Esse comando apresenta uma caixa de diálogo perguntando o intervalo de agrupamento desejado, como mostrado na figura.

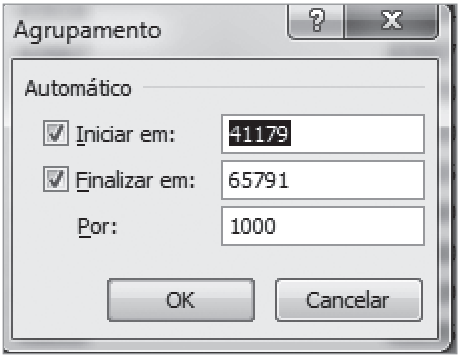


Figura 1.14 – Tabela dinâmica: Agrupamento.

Por exemplo, escolhendo-se o menor valor numérico do Ibovespa para início e o maior valor para término, pode-se agrupar os dados em um intervalo de mil, como mostrado na Figura 1.14. Na caixa **Valores** da tabela dinâmica, deve-se trocar a alternativa selecionada **soma de** para **contagem**. Ao clicar em **soma de**, deve-se escolher a opção **Configuração do Campo de Valor**. Dentro dessa opção, aparece a caixa de **Resumir Valores por**, com diversas opções, como mostrado na figura a seguir. Como o intuito é criar a tabela de classes, deve-se verificar a frequência por faixa, conforme já explicado. Por isso, é preciso escolher a contagem de valor para computar a frequência com que dados do Ibovespa aparecem em cada intervalo.

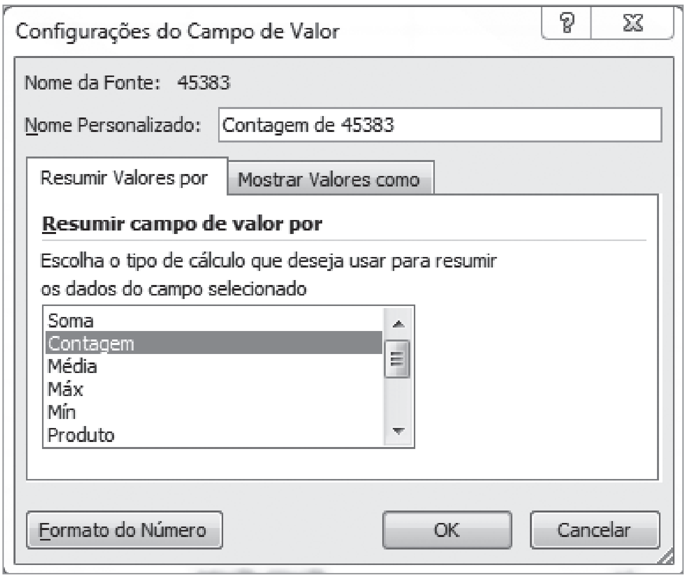


Figura 1.15 – Mudança da tabela dinâmica para frequência de valores nos intervalos.

Desse modo, para o intervalo de mil dados, é possível encontrar a classe modal, ou seja, aquela na qual se encontram mais dados. No caso deste exemplo, a tabela resultante seria assim:

Tabela 1.9 – Dados do Ibovespa para classes com intervalo de mil dados

41.179-42.178	**2**
42.179-43.178	14
43.179-44.178	15
44.179-45.178	12
45.179-46.178	16
46.179-47.178	8
47.179-48.178	3
48.179-49.178	9
49.179-50.178	7
50.179-51.178	10
51.179-52.178	13
52.179-53.178	18
53.179-54.178	13
54.179-55.178	17
55.179-56.178	9
56.179-57.178	4
57.179-58.178	10
58.179-59.178	3
59.179-60.178	4
60.179-61.178	8
61.179-62.178	8
62.179-63.178	15
63.179-64.178	11
64.179-65.178	11
65.179-66.178	4

É fácil verificar que a classe modal é a classe dos pontos do Ibovespa no intervalo entre 52.179 e 53.178. A quantidade de dados presentes nesse intervalo é 18, maior que em outros intervalos. Com essa tabela resultante da tabela dinâmica, é possível fazer um histograma, como visto antes, selecionando as colunas dos intervalos e das frequências. O resultado está na Figura 1.16, que mostra que o intervalo de mil dados tornou o histograma multimodal, com diversos intervalos de valores frequentes muito próximos.

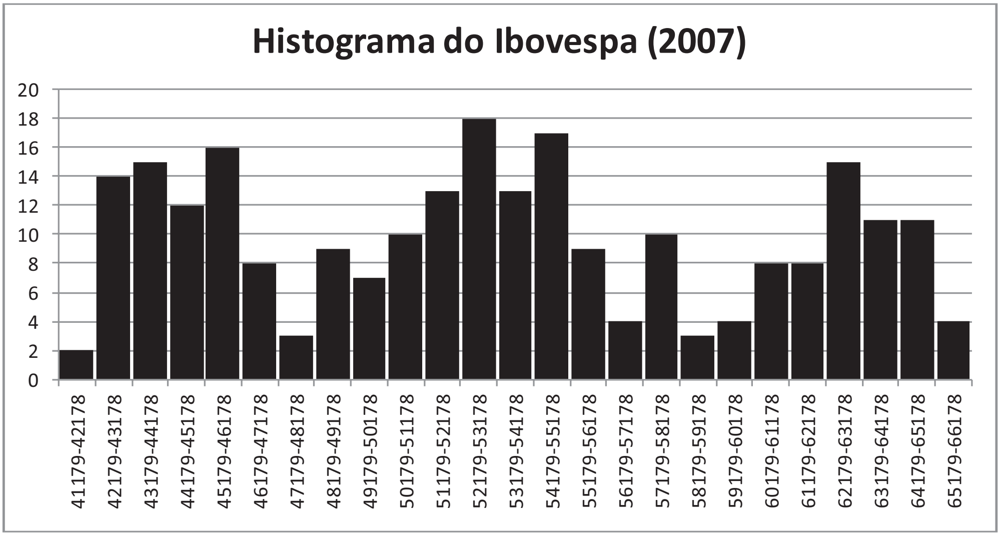


Figura 1.16 – Histograma para tabela de frequências.

1.5 MEDIDAS DESCRITIVAS DE DISPERSÃO NO EXCEL

Assim como as medidas de posição estão programadas no Excel, também as medidas de variabilidade ou dispersão encontram-se na forma de função nas planilhas do Microsoft Excel.

Para estudar a variabilidade de preços, ativos, vendas ou qualquer outra atividade financeira, torna-se muito mais interessante, em vez de calcular as medidas de dispersão sobre os próprios dados, utilizar o retorno (DANÍELSSON, 2011). Para o exemplo do Ibovespa (2007), o retorno positivo (diferença entre fechamento no dia *t* e o dia *t – 1*) indica alta no índice. Valores negativos indicam perdas no índice. O mesmo conceito é aplicado para preços, resultados de vendas e todas as medidas financeiras de mercado.

A Figura 1.17 mostra o retorno dos resultados diários dos fechamentos do Ibovespa. Em especial, a área circulada e destacada no gráfico mostra como a variabilidade aumentou no segundo semestre de 2007. O mercado financeiro estava agitado com dados sobre vendas aquecidas de residências e altíssimos preços de hipotecas nos Estados Unidos. Isso, por sinal, seria uma das principais causas, um ano depois, da quebra do Banco Lehman Brothers.

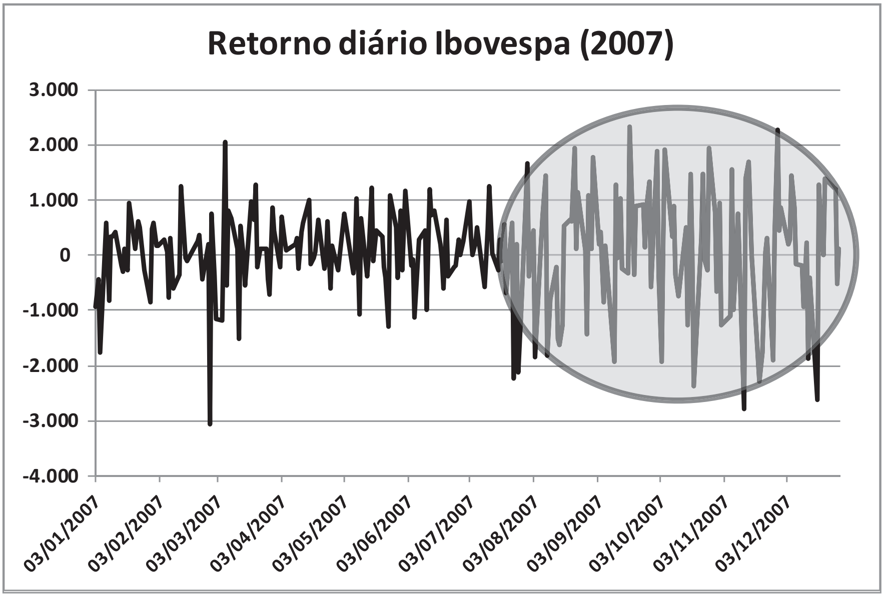


Figura 1.17 – Retorno do Ibovespa em 2007.

Na seção 1.3, foi apresentada a fórmula de cálculo do desvio-padrão, que é o principal indicativo e medida de variabiliade dos dados. No caso do Excel, há a função conhecida como **DesvpadP( )** ou **Desvpad.P( )**, dependendo da versão. Outras variações no nome da função podem aparecer, mas o resultado é sempre o cálculo do desvio-padrão populacional.

O desvio-padrão dos retornos do Ibovespa de 2007 é, supondo que os cálculos estejam na coluna C, programado no Excel conforme a Figura 1.18.

F2 f_x =DESVPAD.P(C2:C245)

	A	B	C	D	E	F
1	02/01/2007	45383				
2	03/01/2007	44445	-938		*Desvio-padrão*	931,7252
3	04/01/2007	44020	-425			

Figura 1.18 – Desvio-padrão dos retornos do Ibovespa em 2007.

O valor do desvio-padrão obtido para os dados desse exemplo é 931,72. Quanto isso significa? Qual a variabilidade desses dados? Para aumentar a “sensibilidade” e a compreensão qualitativa do desvio-padrão, pode-se usar o coeficiente de variação. Como visto na subseção 1.3.2, toma-se o desvio-padrão e divide-se pela média, transformando o resultado em porcentagem.

	A	B	C	D	E	F
1	02/01/2007	45383			*Média*	75,83197
2	03/01/2007	44445	-938		*Desvio-padrão*	931,7252
3	04/01/2007	44020	-425		*Coeficiente de variação*	1228,7%

Figura 1.19 – Coeficiente de variação para o Ibovespa (2007).

O resultado numérico confirma a observação feita do gráfico, ou seja, uma excessiva variabilidade no mercado financeiro de 1.228% sobre os retornos do índice da bolsa de valores. Uma melhor representação para comparar variabilidade dos dados com retorno é a divisão do desvio-padrão dos retornos pela média do índice Ibovespa e não pela média do retorno, como realizado antes.

Quando se compara os desvios dos retornos pelo valor médio do Ibovespa, em vez de comparar pela média dos desvios, que é muito baixa, pode-se encontrar algo mais palpável, com uma variabilidade de 1,75% ao dia no Ibovespa. Nesse caso, a média do Ibovespa foi de 53.145,79 pontos.

Média	53.145,79
Desvio-padrão	931,7252
Coeficiente de variação	1,75%

Figura 1.20 – Média do Ibovespa.

Outra forma de mostrar a variabilidade diária é não calcular apenas o coeficiente de variação estático e anual, pois não é muito sensível a variações e oscilações do mercado financeiro. Pode-se, então, modificar os cálculos para janela de dias úteis, por exemplo, tomando média, desvio-padrão e coeficiente de variação a cada cinco dias. Como se comportaria a variabilidade semanalmente em termos de porcentagem de oscilação?

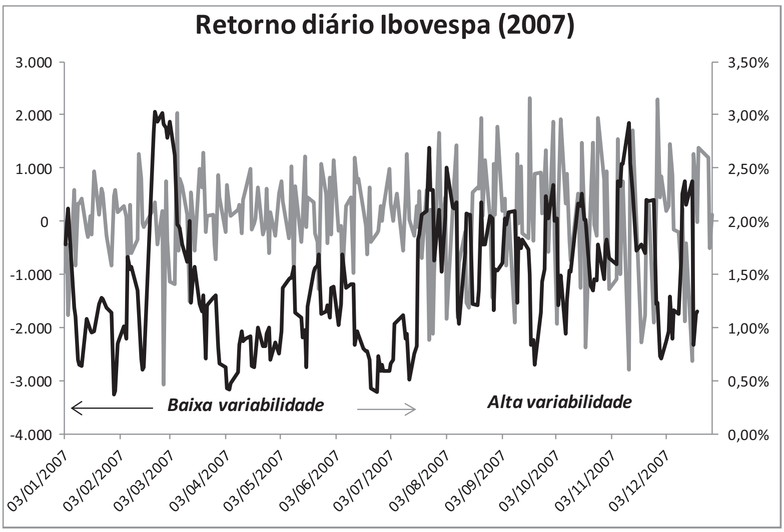


Figura 1.21 – A dinâmica do coeficiente de variação para janela de cinco dias.

A figura apresenta os dados iniciais dos retornos do Ibovespa em cinza e os dados do coeficiente de variação calculado a cada cinco dias em preto. Percebe-se a dinâmica do comportamento da variabilidade, com dominância de baixa variabilidade na oscilação do retorno no primeiro semestre e aumento significativo no segundo semestre de 2007. O coeficiente de variação é apresentado no eixo vertical da direita, com a trajetória em preto. Enquanto no primeiro semestre de 2007 a oscilação foi de 0,5% a 1,5%, no segundo o coeficiente de variação mudou significativamente para valores entre 1% e 3%. Interessante também é ver a comparação da variabilidade entre os dois anos nos retornos do Ibovespa. Para isso, basta fazer os mesmos cálculos dinâmicos para o coeficiente de variação, com janela de cinco dias.

A Figura 1.22, a seguir, mostra quanto a variablidade foi maior para o Ibovespa no ano de 2008 em relação a 2007. A linha cinza no gráfico mostra o coeficiente de variação para os retornos do Ibovespa em 2007, enquanto a linha tracejada mais escura mostra o mesmo coeficiente em 2008. A barra vertical no centro apresenta a separação entre o primeiro e o segundo semestres.

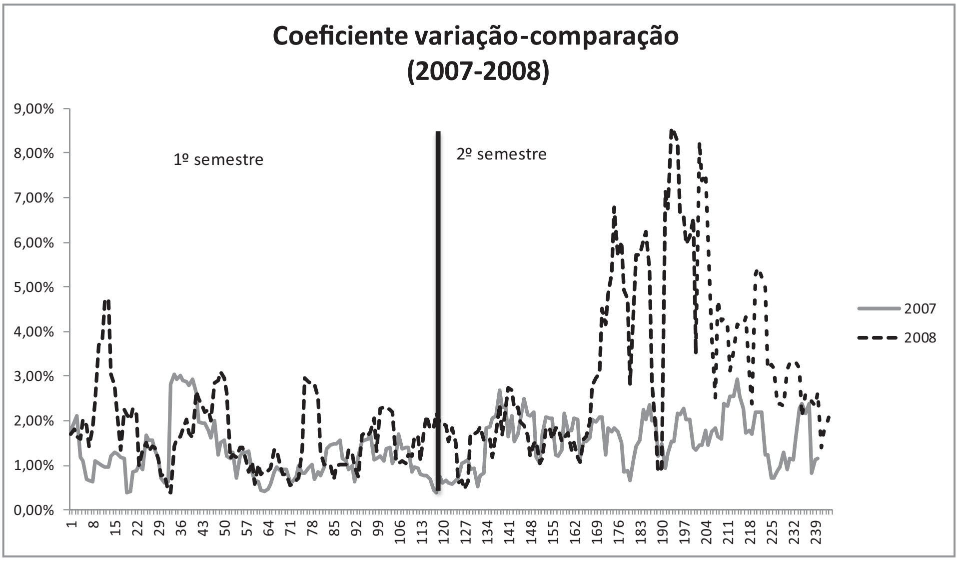
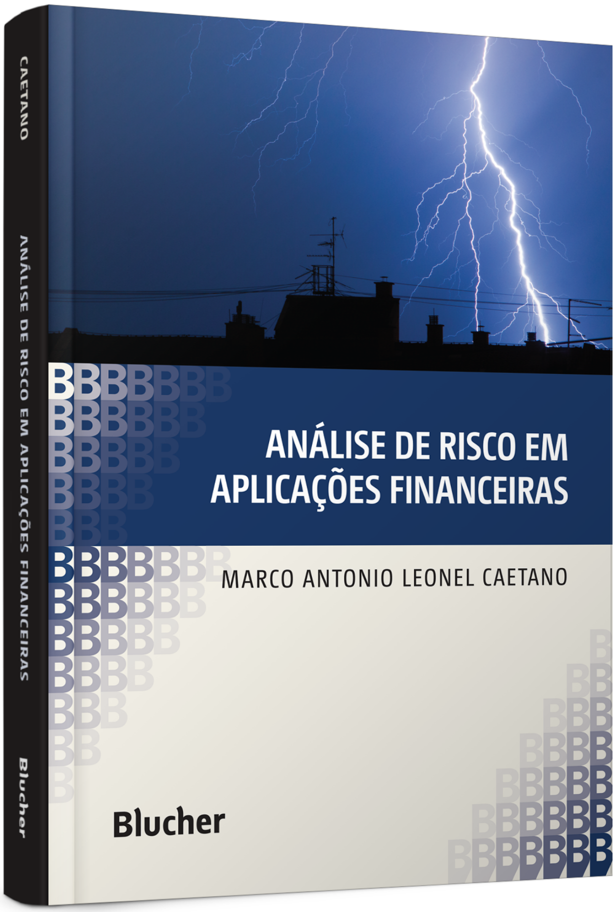


Figura 1.22 – Comparação do coeficiente de variação dinâmico (2007-2008).

Enquanto o máximo valor de variabilidade em 2007 foi de 3%, a variabilidade em 2008 atingiu, na época da crise e das falências nos Estados Unidos, o ápice de 9% de coeficiente de variação.

Essa maneira de mostrar a variabilidade ou a dispersão é mais interessante que a estática para o conjunto todo de dados. Isso se dá porque consegue separar períodos de calmaria de períodos de crise. Procedimento similar pode ser feito para vendas, compras, projeções de resultados etc., pois é interessante para o investidor ou empreendedor ter um entendimento do processo separado por fases ou sazonalidades decorrentes de perturbações financeiras no ano.



Análise de Risco em Aplicações Financeiras

Marco Antonio Leonel Caetano

ISBN: 9788521211440
Páginas: 264
Formato: 17x24 cm
Ano de Publicação: 2017