MODA ESTATÍSTICA: UMA MEDIDA DE TENDÊNCIA CENTRAL

MODA ESTADÍSTICA: UNA MEDIDA DE TENDENCIA CENTRAL

Ancilla Dall'Onder Zatt*

* - Doutoranda do Programa de Pós-Graduação em Educação da Unisinos. ancila@italnet.com.br

Resumo

Este trabalho visa uma reflexão sobre a base conceitual da moda estatística, suas origens, processos de determinação, suas aplicações e processo de aprendizagem. A moda é uma estatística descritiva que indica o valor que mais se repete num conjunto de valores. Karl Pearson utilizou esse termo pela primeira vez em 1895 influenciado pela expressão "estar na moda" quando alguma coisa era frequente. Pode ser obtida pela simples observação dos dados de um conjunto, pelo ponto médio da classe modal de uma distribuição, ou por processos mais elaborados – gráfico e fórmula – de Czuber e de King e, ainda, pela relação empírica de Pearson. Esta faz uso das medidas de tendência central: média, mediana e moda, aplicadas no cálculo da assimetria e da curtose. Cada uma das medidas de tendência central fornece uma visão parcial dos dados, por isso o pesquisador precisa verificar se o parâmetro moda é adequado ao objetivo de seu estudo investigativo. A moda caracteriza-se por sua aplicabilidade a todos os níveis de medida, especialmente aos dados categóricos. É um processo de aprendizagem muito rico ao envolver conceitos prévios e habilidades que se harmonizam num conjunto de relações constitutivas da moda estatística.

Palavras-chave: Moda Estatística, Tendência Central, Aprendizagem.

Resumen

Este trabajo tiene como objetivo refelexionar sobre la base conceptual de la moda estadística, sus orígenes, los procedimientos de determinación, sus aplicaciones y el proceso de aprendizaje. La moda es una estadística descriptiva que indica el valor que más se repite en un conjunto de valores. Karl Pearson utilizó este término por la primera vez en 1895 influenciado por la expresión "estar a la moda" cuando alguna cosa era frecuente. Se puede obtener simplemente por la simple observación de los datos de un conjunto, por el punto médio de la clase modal de una distribución o por procedimeintos más elaborados – gráfico y fórmula – de Czuber y King y también por la relación empírica de Pearson. Esta hace uso de medidas de la tendencia central: media, mediana y moda, utilizadas para calcular la asimetría y la curtosis. Cada una de las medidas de tendencia central proporciona una visión parcial de los datos, por lo que el investigador debe verificar si el parámetro moda es adecuado para los fines de su estúdio de investigación. La moda se caracteriza por su aplicabilidad a todos los niveles de medición, sobre todo para datos categóricos. Es un proceso de aprendizaje muy rico, al envolver conceptos prévios y habilidades que se armonizan en un conjunto de relaciones constitutivas de la moda estadística.

Palabras claves: Moda Estadística, Tendencia Central, Aprendizaje.

Introdução

Em Estatística denominam-se medidas de tendência central: a média, a mediana e a moda. Para Triola (1999, p. 31): "Uma medida de tendência central é um valor no centro ou no meio de um conjunto de dados".

Essas medidas dependem da definição de centro de um conjunto de valores ou de uma distribuição que pode ser interpretado de várias maneiras (BERQUÓ *et al*., 1981). Podem indicar o salário esperado que um trabalhador declara ao ser entrevistado, o salário mais frequente na empresa ou, ainda, o valor salarial central abaixo do qual/acima do qual está situada metade de todos os salários pagos pela empresa.

A moda, objeto deste artigo, "é o valor que ocorre com maior frequência num conjunto de dados, isto é, o valor mais comum" (SPIEGEL, 1976, p. 74). A palavra "moda" significa, no cotidiano, ser "muito usado" e segundo Clegg (1995) expressa com propriedade o significado da moda estatística. Esta é o valor que se repete o maior número de vezes, num conjunto de valores, isto é, o mais frequente.

São muito comuns expressões que mencionam a preferência por determinado produto, maior audiência entre emissoras, obras mais vendidas, candidato mais votado, numeração de calçados de maior procura e outras que passam a ideia de um valor mais frequente, estatisticamente denominado moda.

Convém assinalar que expressões como: "...a maioria dos sistemas de ensino brasileiro não havia adotado o Ensino Fundamental com duração de nove anos" (BRANDÃO; PASCHOAL, 2009), num conjunto de dados "a maioria" nem sempre representa a moda estatística (HOUT, 1999).

As obras de Estatística são unânimes em referenciar o parâmetro moda no cálculo das medidas de assimetria e curtose, mas poucas oferecem seu tratamento estatístico.

A referência mais remota que se tem do uso da moda, citada por Wallis e Roberts em sua obra *Curso de Estatística*, é o que se refere ao cerco dos plateus pelos peloponésios em 428 a.C. No inverno desse ano os plateus, juntamente com os atenienses, estavam sitiados pelos peloponésios e pelos beócios. Armaram um plano para escaparem forçando a passagem pelas muralhas inimigas. Para construir escadas que alcançassem a altura da muralha inimiga muitas pessoas contaram e recontaram, ao mesmo tempo, as camadas de tijolos. Ainda que alguém errasse a contagem, a maioria haveria de ter acertado. Foi dessa forma que obtiveram o comprimento necessário para as escadas alcançarem o objetivo.

O termo moda foi utilizado pela primeira vez por Karl Pearson, em 1895, influenciado pela maneira de falar das pessoas ao afirmarem que tal objeto está na moda, com o significado de coisa mais frequente (GONÇALVES, 1978). Essa definição permite observar que um conjunto de valores pode possuir mais de uma moda.

Diz-se que um conjunto é unimodal, bimodal, trimodal ou plurimodal, de acordo com o número de modas que apresenta. A ausência de uma moda caracteriza o conjunto como amodal. Encontra-se na estrutura das palavras, o conhecimento prévio para a construção conceitual da classificação dos conjuntos de valores em função da presença ou ausência da moda.

Desvendando os processos de determinação da moda

Sendo a moda de um conjunto de dados o valor que mais se repete (BISQUERA *et al.*, 2004) parece simples a sua identificação. Segundo Stevenson (1982, p. 23): "A moda funciona como medida descritiva quando se trata de contar dados". No caso de uma distribuição de frequência a classe modal será a que apresenta maior frequência (20 ⊢ 25), no exemplo citado, pois contém o valor da moda na distribuição. O ponto médio representativo da classe modal é denominado moda bruta que no exemplo dado é 22,5 peças diárias. Numa distribuição de frequência a classe modal nem sempre corresponde à classe que contém a mediana.

Karl Pearson observou a existência de uma relação empírica que permite calcular a moda quando são conhecidas a média ($\bar{x}$)e a mediana (Me) de uma distribuição moderadamente assimétrica. Essas condições satisfazem a relação empírica Mo = 3 Me – 2$\bar{x}$ que, no exemplo apresentado, corresponde a Mo = 3(23,08) – 2(23), então Mo = 23,24 peças. Esse processo supõe o domínio conceitual simetria/assimetria, empírica, o cálculo da média aritmética e da mediana. Em uma distribuição simétrica as três medidas – $\bar{x}$, Me, Mo – são exatamente iguais.

Convém lembrar que a fórmula de Pearson pode ser empregada com bons resultados quando os valores da média e da mediana forem conhecidos e a distribuição não for muito simétrica.

Considere-se a distribuição das vendas diárias do setor de peças de uma determinada loja para explicitar os processos de cálculo da moda indicados.

Nº de peças (X)	Nº de dias (fi)	Xi	fac	XiFi
5 ⊢ 10	3	7,5	3	22,5
10 ⊢ 15	9	12,5	12	112,5
15 ⊢ 20	12	17,5	24	210,0
20 ⊢ 25	26	22,5	50	585,0
25 ⊢ 30	15	27,5	65	412,5
30 ⊢ 35	13	32,5	78	422,5
35 ⊢ 40	2	37,5	80	75,0
Total	80	-	-	1.840,0

Moda bruta = 22,5 peças

$$\overline{X} = \frac{\Sigma Xifi}{\Sigma fi} \qquad \overline{X} = \frac{1840}{80} \qquad \overline{X} = 23 \text{ peças}$$

Média = $\overline{X}$

ΣXifi = somatório dos produtos frequência por ponto médio classe a classe

Σfi = somatório da fi.(total)

$$Me = li + \frac{Po - fai}{fi} \cdot h \qquad \text{onde} \qquad Po = \frac{\Sigma fi}{2} \qquad Po = 40$$

$$Me = 20 + \frac{40 - 24}{26} \cdot 5 \qquad Me = 20 + 3{,}076 \qquad Me \cong 23{,}08$$

Para a mediana:

Me = mediana

li = limite inferior da classe de localização da Me

Po = localização da classe da Me

fai = frequência acumulada classe inferior à da Me

fi = frequência simples da classe da Me

h = intervalo de classe

Moda de Pearson = 3(Me) – 2($\bar{x}$)

Moda = 3(23,08) – 2(23)

Moda ≅ 23,24 peças

Desejando-se obter a moda com mais exatidão, empregam-se os processos de Czuber e King, os quais apresentam possibilidade de determinação gráfica e um raciocínio matemático em suas formulações. Czuber desenvolve uma forma mais aproximada para o cálculo da moda partindo de um processo gráfico.

Para determinar graficamente a moda Czuber parte do histograma (Figura 1), utilizando os três retângulos correspondentes à classe modal e às classes adjacentes. A moda será o valor do limite inferior da classe modal acrescida de um valor "X" determinado pela intersecção dos segmentos $\overline{AB}$ (que une o limite superior da classe que antecede a classe modal ao limite superior da classe modal) e $\overline{CD}$ (que une o limite inferior da classe modal ao inferior da classe posterior à modal). Portanto: Mo = li + X.

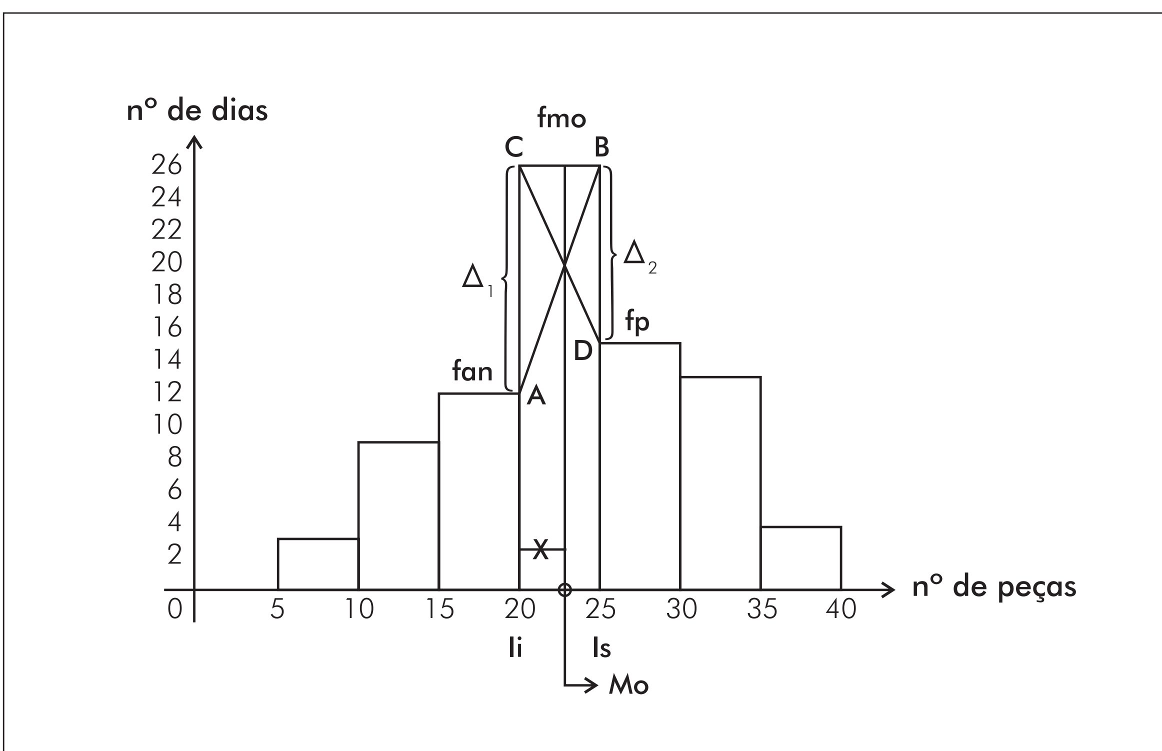

Figura 1. Processo gráfico de Czuber para determinação da moda.

A observação da figura do histograma mostra o uso dos conceitos prévios de semelhança entre os triângulos, de proporcionalidade e a hipótese de Czuber: “A moda divide o intervalo da classe modal em distâncias proporcionais às diferenças entre a frequência da classe modal com a frequência das classes adjacentes”.

$$\frac{X}{h - X} = \frac{\Delta_1}{\Delta_2}$$

Então:

$$\Delta_2 . X = \Delta_1 (h - X)$$
$$\Delta_2 . X = \Delta_1 h - \Delta_1 X$$
$$\Delta_1 X + \Delta_2 X = \Delta_1 h$$
$$X(\Delta_1 + \Delta_2) = \Delta_1 h$$

Fazendo-se:
fmo = frequência modal
fan = frequência anterior à modal
fp = frequência posterior à modal
h -= intervalo de classe
li = limite inferior da classe modal
Mo = moda

$$X = \frac{\Delta_1 h}{\Delta_1 + \Delta_2} \quad \text{ou} \quad X = \frac{\Delta_1}{\Delta_1 + \Delta_2} . h$$

e sabendo que: $\begin{cases} \Delta_1 = fmo - fan \\ \Delta_2 = fmo - fp \end{cases}$ substitui-se pelas equivalências

então $$X = \frac{fmo - fan}{fmo - fan + fmo - fp} . h$$

$$X = \frac{fmo - fan}{2fmo - (fan + fp)} . h$$

e como Mo = li + x, substituindo-se X pelo seu valor, tem-se a fórmula de Czuber:

$$Mo = li + \frac{fmo - fan}{2fmo - (fan + fp)} . h$$

O processo gráfico de Czuber embasa o processo matemático, e sua construção formal requer o conhecimento de conceitos prévios de geometria, proporcionalidade e fatoração.

No exemplo da distribuição da frequência das peças vendidas, a moda de Czuber indica:

$$Mo = 20 + \frac{26-12}{52-(12+15)} \cdot 5$$

$$Mo = 20 + \frac{14}{52-27} \cdot 5$$

$$Mo = 20 + \frac{70}{25}$$

$$Mo = 20 + 2{,}8 \qquad Mo = 22{,}8 \text{ peças}$$

O estudante familiarizado com o cálculo pode encontrar a fórmula de Czuber através da parábola construída de modo a passar pelos pontos médios da classe modal e das classes adjacentes a ela.

O processo de W. I. King apresenta sua forma geométrica de determinação através do histograma conforme a Figura 2.

Percebe-se que a Mo = li + X.

Traçando, na figura, em continuidade ao segmento do limite superior da classe modal, a projeção do limite inferior da classe modal, tem-se $\overline{AB}$ que faz intersecção com o eixo da abscissa – escala numerada – onde se lê a moda das peças vendidas.

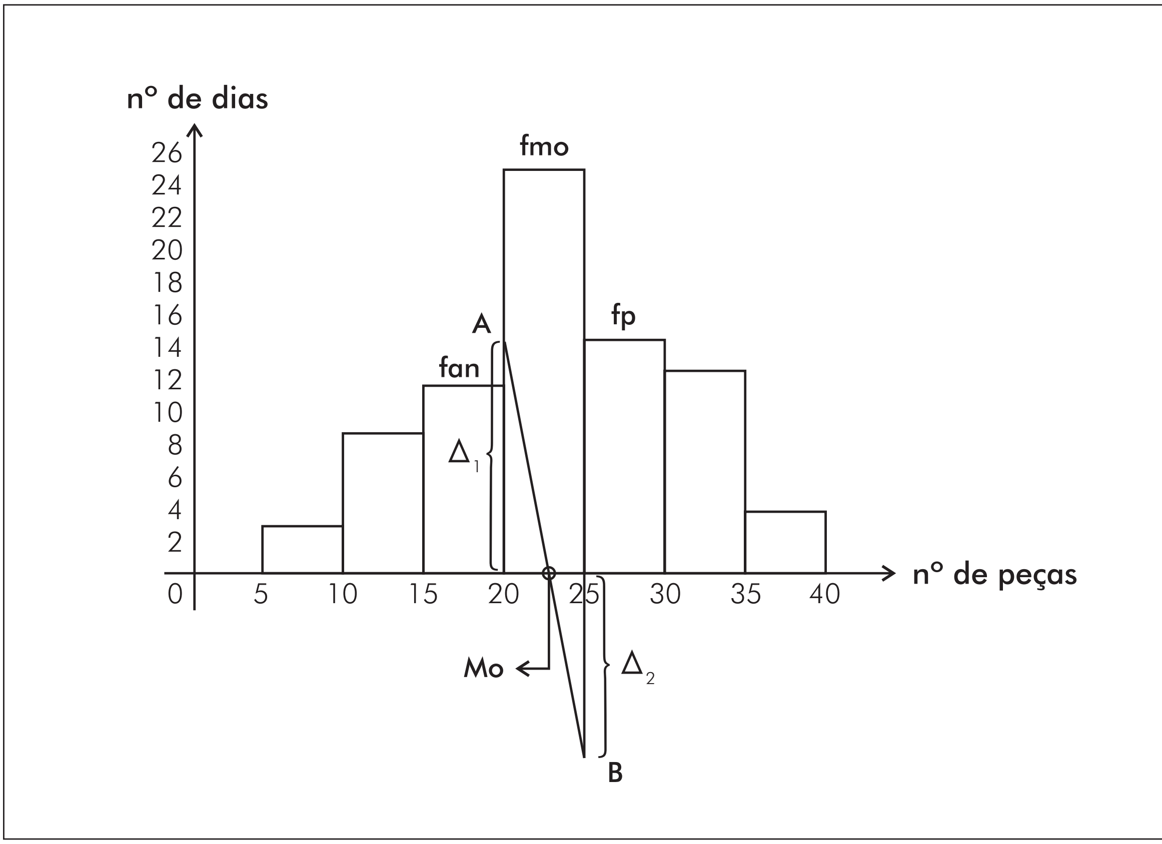

Figura 2. Processo gráfico de King para determinação da moda.

A proposta de King pouco difere da de Czuber baseada nos conceitos de semelhança entre os triângulos e proporcionalidade. Retoma-se a ideia de a moda ser equivalente ao valor do limite inferior da classe modal acrescido de um valor "X" correspondente ao segmento entre o limite inferior da classe modal e o ponto de intersecção com o eixo da abscissa.

O processo de cálculo baseia-se na proporcionalidade pela semelhança dos triângulos de acordo com a figura acima e na hipótese de King: "A moda divide o intervalo da classe modal em distâncias inversamente proporcionais às frequências das classes adjacentes".

$$\frac{X}{h - X} = \frac{\Delta_1}{\Delta_2} \qquad \text{sendo que: } \begin{cases} \Delta_1 = fp \\ \Delta_2 = fan \end{cases}$$

fp = frequência da classe posterior à modal

fan = frequência da classe anterior à modal

$$X.\ \Delta_2 = \Delta_1(h - X)$$
$$X\Delta_2 = \Delta_1 h - \Delta_1 X$$
$$\Delta_1 X + \Delta_2 X = \Delta_1 h$$
$$X(\Delta_1 + \Delta_2) = \Delta_1 h$$
$$X = \frac{\Delta_1}{\Delta_2 + \Delta_2}\ .\ h$$

$$\text{então:}\quad X = \frac{fp}{fp + fan}\ .\ h$$

e substituindo-se "X" na relação inicial Mo = li + X, temos que:

$$Mo = li \ + \frac{fp}{fan + fp}\ .\ h$$

Para a dedução desta fórmula para dados agrupados em classes de frequência o autor utilizou os mesmos conceitos prévios do processo de Czuber. A moda de King para o exemplo das peças vendidas indica:

$$Mo = 20 + \frac{15}{12 + 15}\ .\ 5$$

$$Mo = 20 + \frac{15}{27}\ .\ 5$$

$$Mo \cong 20 + 2{,}77$$

$$Mo \cong 22{,}77 \text{ peças}$$

As propostas de Czuber e King para a moda elaborada apresentam certa similaridade em seu raciocínio e diferem no que se refere às frequências. King baseia-se na influência das frequências adjacentes sobre a classe modal, e Czuber leva em consideração não apenas as frequências das classes adjacentes, mas também da frequência da classe modal, segundo Madsen Barbosa (p. 99). Observa-se que os valores extremos não afetam o cálculo da mesma.

Foi possível verificar nos exemplos dados que o cálculo da moda não apresentou os mesmos resultados, o que geralmente ocorre em virtude do processo adotado.

Considerações finais

A moda faz parte das medidas de tendência central – média, mediana e moda – utilizadas na análise da assimetria/simetria e curtose. É um parâmetro fácil de calcular e não é afetada pelos valores extremos, mas seu valor é fortemente afetado pela maneira como as classes são constituídas. Segundo Bunchaft e Kellner (1998, p. 119): "A moda tem como característica importante a sua aplicabilidade a todos os níveis de medida – nominal, ordinal e intervalar – sendo seu emprego desejável em se tratando de dados em categorias, ou seja, distribuições de variáveis qualitativas". Essa ideia é sustentada também por Huot (1999) ao afirmar que a moda é especialmente útil para reduzir a informação de um conjunto de dados qualitativos, apresentados em forma de nomes ou em categorias.

Entre os muitos usos da moda estatística está a pesquisa quando se realiza pesquisa descritiva ou quando se trata de dados qualitativos como é o caso da investigação de Del Pino e Porto (2008). No ramo de confecções, indica os tamanhos mais usuais, ao calçadista, o número ou numeração de calçados mais vendidos. É ainda a moda que indica o salário predominante na empresa ou a predominância de evasão na escola.

Bunchaft e Kellner (1998) e Callegari-Jacques (2003) afirmam que para descrever distribuições bimodais é interessante identificar as duas modas, pois além de acentuadas podem evidenciar características que a média e a mediana não podem descrever. Lapponi (1997) aponta entre as desvantagens da moda, não usar todos os dados disponíveis, por estar afastada do centro das observações. A moda nada acrescenta na descrição dos dados, quando todos ou quase todos os valores ocorrem com frequência aproximada (STEVENSON, 1981).

Ao pesquisador cabe a tarefa de identificar em que situação deve usar a moda como medida descritiva ou na pesquisa de natureza qualitativa. Sugere-se ao professor no ensino explorar as ricas possibilidades conceituais prévias e em construção na aprendizagem da moda, em suas relações interdisciplinares.

Referências

BERQUÓ, Elza Salvatori *et al*. *Bioestadística*. 1. ed. rev. São Paulo: EPU, 1981.

BISQUERA, Rafael *et. al*. *Introdução à estatística*: enfoque informático com o pacote estatístico SPSS. Porto Alegre: Artmed, 2004.

BRANDÃO, Carlos da Fonseca; PASCHOAL, Jaqueline Delgado (Org.). *Ensino Fundamental de nove anos*. Teoria e prática na sala de aula. São Paulo: Avercamp, 2009.

BUNCHAFT, Guenia; KELLNER, Sheilah R. Oliveira. *Estatística sem mistérios*. V. I. 2. ed. Petrópolis: Vozes, 1997.

CALLEGARI-JACQUES, Sídia M. *Bioestatística*. Porto Alegre: Artmed, 2003.

CLEGG, Frances. *Estatística para todos*. Lisboa: Gradiva, 1995.

DEL PINO, Augusto Burkert; PORTO, Gilvane Cateano. Exclusão escolar: a história continua no século XXI. *Educação Unisinos*, São Leopoldo, v. 12, n. 2, p. 100-110, maio/ago. 2008.

GONÇALVES, Fernando A. *Estatística descritiva*. 2. ed. São Paulo: Atlas, 1978.

HUOT, Rijean. *Métodos quantitativos para as Ciências Humanas*. Lisboa: Piaget, 1999.

LAPPONI, Juan Carlos. *Estatística usando Excel 5 e 7.* São Paulo: Lapponi Treinamento e Editora, 1997.

MADSEN BARBOSA, Ruy. *Estatística elementar*. V. 1. São Paulo: Benetti, s.d.

SPIEGEL, Murray R. *Estatística*. ed. rev. São Paulo: McGraw-Hill do Brasil, 1976.

STEVENSON, William. *Estatística aplicada à administração*. São Paulo: Harper & Row do Brasil, 1981.

TRIOLA, Mário F. *Introdução à Estatística*. 7 ed. Rio de Janeiro: LTC, 1999.

WALLIS, W. Allen; ROBERT, Hary V. *Curso de Estatística*. V. 1. Rio de Janeiro: Fundo de Cultura, s.d.