

# UMA COMPARAÇÃO ENTRE ANÁLISE DE VARIÂNCIA E ANÁLISE DE REGRESSÃO APLICADAS AO MERCADO DE AÇÕES

1. Introdução;
2. O modelo de análise adotado por Figueiredo;
3. Comparação entre análise de variância e análise de regressão;
4. Resultados e conclusões.

Denisard C. de Oliveira Alves \*

\* Professor do Departamento de Economia e do Instituto de Pesquisas Econômicas da USP.

## 1. INTRODUÇÃO

O objetivo deste trabalho é a comparação de duas técnicas econométricas para estudar sua utilidade na análise do mercado de ações. Com o desenvolvimento do mercado de ações surgiu a necessidade de elaboração de indicadores técnicos de análise e previsão do comportamento do mercado a fim de orientar os investidores na formação de seus portfólios. No final de 1970, as principais bolsas do país passaram a divulgar uma série de índices, com base em dados colhidos dos balanços das empresas e dos pregões diários.

Com o objetivo de testar a validade desses índices, Figueiredo,<sup>1</sup> realizou uma análise estatística de dois períodos distintos, de 12 meses, caracterizados como um período de alta no mercado (agosto de 1970 a julho de 1971) e um período de baixa (junho de 1971 a maio de 1972). Figueiredo usou a técnica estatística de análise de variância. Neste trabalho nós iremos refazer o trabalho de Figueiredo com base nos mesmos dados por ele utilizados,<sup>2</sup> mas usando análise de regressão ao invés de análise de variância. Na primeira parte deste trabalho o modelo de análise adotado por Figueiredo é sucintamente explicado. Na segunda parte é feita uma comparação entre a técnica utilizada por Figueiredo e a técnica de análise de regressão. Finalmente, na última parte, os resultados das duas técnicas são analisados e as conclusões do trabalho são apresentadas.

## 2. O MODELO DE ANÁLISE ADOTADO POR FIGUEIREDO

### 2.1 Procedimento

Para cada um dos períodos analisados ele adotou o seguinte procedimento:

2.1.1 Sete grupos, de 10 ações cada um, foram selecionados no início de cada período de análise.

Estes grupos foram separados da seguinte forma:

- a) seis grupos utilizando seis diferentes índices técnicos, conforme será explicado a seguir;
- b) Um grupo de escolha inteiramente aleatória.

2.1.2 Ao fim de cada período, foi computada a lucratividade (período de alta) ou perda (período de baixa) para cada um dos sete grupos selecionados. A lucratividade ou perda foi medida como uma porcentagem do preço de aquisição da ação no início de cada período.

2.1.3 Com base nos dados apresentados, foi realizada uma análise de variância para verificar se as médias de lucratividade ou de perda, de cada um dos grupos selecionados, apresentavam diferenças significativas ou não.

## 2.2 Índices utilizados (período de alta)

Para o período de alta (agosto de 1970 a julho de 1971), foram utilizados os seguintes índices técnicos (as amostras correspondentes são apresentadas no anexo 1 do trabalho de Figueiredo).

**Grupo A: Ações de empresas com maior valor patrimonial por ação**

O valor patrimonial é obtido dividindo-se a soma do capital e das reservas livres pelo número de ações da empresa.

**Grupo B: Ações de maior índice preço/lucro (P/L)**

O índice preço/lucro representa a divisão do preço da ação na época da análise (agosto de 1970) pelo lucro por ação do último exercício.

**Grupo C: Ações de menor índice preço/lucro**

**Grupo D: Ações de maior negociabilidade nos últimos 12 meses**

A negociabilidade é medida pela frequência com que a ação é transacionada nos pregões da bolsa.

**Grupo E: Ações que apresentavam maior quociente:**  $\frac{\text{Valor patrimonial}}{\text{Índice preço/lucro}}$

Este índice composto é obtido da seguinte forma:

$VP = \frac{A}{N}$ , onde  $A$  é patrimônio da empresa e  $N$  é o número de ações.

$P/L = \frac{P}{L/N}$ , onde  $P$  é o preço da ação,  $L$  o lucro por ação e  $N$  o número de ações.

$$\frac{VP}{P/L} = \frac{A}{N} : \frac{P}{L/N} = \frac{AL}{N^2P}$$

Como se deve observar, quanto maior o patrimônio e o lucro da empresa e quanto menor o número de ações e o preço, tanto melhor seria a ação, do ponto de vista deste índice.

**Grupo F: Ações que apresentaram maior quociente:**  $\frac{\text{negociabilidade} \times \text{valor patrimonial}}{\text{índice preço/lucro}}$

Este é o outro índice de composto obtido acrescentando-se ao anterior (grupo E) o índice de negociabilidade. Se chamamos  $C$  ao índice de negociabilidade, teremos a seguinte expressão para o índice:

$$\frac{CAL}{N^2P}$$

**Grupo G: Ações selecionadas aleatoriamente**

Essas 10 ações foram selecionadas inteiramente ao acaso, utilizando-se uma tabela de números aleatórios.

Em resumo, os resultados obtidos para cada uma das sete amostras são os seguintes:

Tabela 1

| Grupos                           | Lucratividade média |
|----------------------------------|---------------------|
| A = maior valor patrimonial (VP) | 280,21%             |
| B = maior P/L                    | 384,82%             |
| C = menor P/L                    | 680,44%             |
| D = maior negociabilidade        | 388,79%             |
| E = maior $\frac{VP}{V/L}$       | 300,71%             |
| F = maior $\frac{CAL}{N^2P}$     | 402,09%             |
| G = seleção aleatória            | 288,35%             |

59

## 2.3 Índices utilizados (período de baixa)

Para o período de baixa (junho de 1971-maio de 1972), foram utilizados os mesmos índices do período de alta, com exceção do índice relativo ao Grupo F, que deu lugar a um novo índice, apresentado regularmente na *Revista da Bolsa de Valores de São Paulo*, com o nome de índice de giro. Este índice mostra a participação percentual de um determinado tipo de ação relativamente ao total de ações do mesmo tipo, de uma empresa.

Assim:

$$\text{Giro} = \frac{\text{quantidade total negociada da ação pref. A da companhia X}}{\text{quantidade total de ações pref. A excedentes da companhia X}}$$

A decisão de trocar o índice  $F$  no segundo período decorreu das seguintes observações:

a) o índice  $\frac{\text{negociabilidade x valor patrimonial}}{\text{índice preço/lucro}}$

é, num certo sentido, redundante, pois se acha representado nos demais índices;

b) o índice de giro só surgiu em 1971 e foi considerada convenientemente sua utilização, pois se tratava de uma nova variável de análise;

c) a manutenção de sete grupos de índices era conveniente do ponto de vista de computação, pois o mesmo programa poderia ser utilizado nos dois casos.

As seguintes perdas médias foram observadas no período de baixa, em cada uma das amostras de 10 ações (as ações componentes de cada uma das amostras são apresentadas no anexo 2).

Tabela 2

| Grupos                      | Perda média |
|-----------------------------|-------------|
| A = maior valor patrimonial | 41,72%      |
| B = maior P/L               | 43,22%      |
| C = menor P/L               | 45,36%      |
| D = maior negociabilidade   | 44,45%      |
| E = maior $\frac{LP}{P/L}$  | 34,92%      |
| F = maior giro              | 43,06%      |
| G = seleção aleatória       | 43,66%      |

60

## 2.4 Análise de variância

Através do modelo de análise de variância Figueiredo testou a seguinte hipótese nula:

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_G = 0$$

Revista de Administração de Empresas

onde  $\alpha_g$  ( $g = 1, 2, \dots, G$ ) é o desvio da média de lucratividade do  $g$ -ésimo grupo relativamente à média global de lucratividade,  $\mu$ .

$H_1 : H_0$  não é verdadeira, ou seja, essa hipótese implica que pelo menos duas das médias, em cada período de análise, não são iguais.<sup>3</sup>

## 3. COMPARAÇÃO ENTRE ANÁLISE DE VARIÂNCIA E ANÁLISE DE REGRESSÃO

Modelos de análise de variância podem ser formulados em termos de análise de regressão pela utilização do método de variáveis auxiliares ou dummies. Os dois enfoques são equivalentes, ou seja, eles descrevem o mesmo fenômeno e levam aos mesmos resultados em termos de teste de hipóteses. Isto pode ser comprovado formalmente. Seja uma variável aleatória  $Y$  normalmente distribuída, cuja média dependa de um grupo de características que se reflete em uma classificação de  $Y$  em  $G$  grupos. A variância de  $Y$  em cada grupo é suposta constante, e as observações são supostas independentes. No modelo de análise de variância nós dividimos todos os valores observados de  $Y$  em  $G$  grupos, de acordo com características dadas, e formulamos o modelo como:

$$a) Y_{gt} = \mu + \alpha_g + \epsilon_{gt}, \quad (t = 1, \dots, n_g; g = 1, 2, \dots, G),$$

onde  $Y_{gt}$  é a  $t$ -ésima observação de  $Y$  no  $g$ -ésimo grupo,  $\mu$  é a média global de  $Y$ ,  $\alpha_g$  é o desvio da média do  $g$ -ésimo grupo relativamente a  $\mu$ , e  $\epsilon_{gt}$  é o termo aleatório.

$$b) \sum_{g=1}^G \alpha_g = 0$$

Vamos agora comparar o modelo de análise de variância com o modelo de análise de regressão correspondente:

$$c) Y_{gt} = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \dots + \beta_G X_{Gt} + \epsilon_{gt}$$

onde  $X_{gt}$  é uma variável dummy que assume o valor 1 se a observação de  $Y$  pertence ao  $g$ -ésimo grupo; e assume o valor zero caso contrário; ( $g = 2, 3, \dots, G$ ).

O modelo (c) é o modelo de análise de regressão com variáveis explicativas qualitativas.  $\beta_1$  é a  $E(Y/X_{2t} = 0, X_{3t} = 0, \dots, X_{Gt} = 0)$  ou seja,  $\beta_1$  é a média populacional do grupo 1 e  $X_{1t}$  é a variável que está na base.<sup>4</sup>

Para demonstrar a equivalência dos dois modelos representados pelas equações (a) e (c) vamos especificar os coeficientes de (c) em termos dos parâmetros de (a) e vice-versa. Na tabela 3 formulamos os dois modelos para cada grupo separadamente.

Tabela 3

| Grupos | Modelo de análise de variância | Modelo de análise de regressão  |
|--------|--------------------------------|---------------------------------|
| 1      | $E(Y_{1t}) = \mu + \alpha_1$   | $E(Y_{1t}) = \beta_1$           |
| 2      | $E(Y_{2t}) = \mu + \alpha_2$   | $E(Y_{2t}) = \beta_1 + \beta_2$ |
| .      | .                              | .                               |
| .      | .                              | .                               |
| .      | .                              | .                               |
| G      | $E(Y_{Gt}) = \mu + \alpha_G$   | $E(Y_{Gt}) = \beta_1 + \beta_G$ |

Ou seja,

$$d) \begin{cases} \mu + \alpha_1 = \beta_1 \\ \mu + \alpha_2 = \beta_1 + \beta_2 \\ \cdot \\ \cdot \\ \mu + \alpha_G = \beta_1 + \beta_G \end{cases}$$

Somando-se essas G equações temos:

$$e) \sum_{g=1}^G \mu + \sum_{g=1}^G \alpha_g = \sum_{g=1}^G \beta_1 + \sum_{g=2}^G \beta_g,$$

ou seja,

$$f) G \cdot \mu + \sum_{g=1}^G \alpha_g = G \cdot \beta_1 + \sum_{g=2}^G \beta_g,$$

ou,

$$g) \mu = \beta_1 + \frac{1}{G} [\beta_2 + \beta_3 + \dots + \beta_G], \text{ pois}$$

$$\text{por (b)} \sum_{g=1}^G \alpha_g = 0$$

$$h) \alpha_1 = -\frac{1}{G} [\beta_2 + \beta_3 + \dots + \beta_G] \text{ por (d) e (g).}$$

$$i) \alpha_2 = \beta_2 - \frac{1}{G} [\beta_2 + \beta_3 + \dots + \beta_G] \text{ por (d) e (g), e assim sucessivamente até}$$

$$j) \alpha_G = \beta_G - \frac{1}{G} [\beta_2 + \beta_3 + \dots + \beta_G].$$

A hipótese a ser testada pelo modelo de análise de variância é que não existe diferença entre as médias de Y nos G grupos; isto é,

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_G = 0$$

$$H_1: H_0 \text{ não é verdadeira}$$

É fácil verificar, a partir das equações (g) a (j), que essa hipótese nula é exatamente equivalente à hipótese de que as declividades do modelo de regressão (a) são conjuntamente iguais a zero, ou seja,

$$H_0: \beta_2 = \beta_3 = \dots = \beta_G = 0$$

$$H_1: H_0 \text{ não é verdadeira}$$

Como é sabido, para testar  $H_0$  no modelo de análise de regressão utiliza-se o teste de  $F^5$ . Esse teste é exatamente equivalente ao teste de  $F$  do modelo de análise de variância dos livros-textos de estatística.<sup>6</sup> Por outro lado, no modelo de regressão é possível testar a hipótese de que qualquer um dos  $\beta$ 's seja individualmente nulo através do teste de  $t$ , enquanto que no modelo de análise de variância isso não é possível. Evidentemente, seria possível comparar as médias duas a duas, através do teste  $t$ .

## 4. RESULTADOS E CONCLUSÕES

### 4.1 Resultados da análise de regressão para o período de alta

Tabela 4  
Variável dependente: lucratividade

| Variáveis independentes | Grupo correspondente  | Coefficiente estimado | Teste "t" |
|-------------------------|-----------------------|-----------------------|-----------|
| C                       | Maior VP              | 280,210               |           |
| X <sub>2</sub>          | Maior P/L             | 74,610                | 0,292     |
| X <sub>3</sub>          | Menor P/L             | 400,230               | 1,565     |
| X <sub>4</sub>          | Maior negociabilidade | 111,880               | 0,438     |
| X <sub>5</sub>          | Maior VP/PL           | 20,500                | 0,080     |
| X <sub>6</sub>          | Maior C.VP/PL         | 125,180               | 0,489     |
| X <sub>7</sub>          | Seleção aleatória     | 8,140                 | 0,032     |

$$R^2 = 0,0533 \quad F = 0,591736 \quad n = 70$$

Soma dos quadrados dos resíduos = 20 598 000

Como foi discutido no item 3 as variáveis independentes são variáveis dummies representando cada um dos seis grupos de índices técnicos; o grupo que está na base é o grupo de firmas com maior VP, ou seja, o valor estimado da lucratividade de média das ações com maior VP é 280,210, que é a estimativa do coeficiente  $\beta_1$  na equação (c) com  $G = 7$ . Vejamos agora o significado dos demais coeficientes estimados da equação (c) apresentados na tabela 4. Nós sabemos que:

$$E(Y_{2t}/X_1 = 0; X_2 = 1; X_3 = 0; X_4 = 0; X_5 = 0; X_6 = 0; X_7 = 0) = \beta_1 + \beta_2,$$

ou seja,  $\hat{\beta}_1 + \hat{\beta}_2$ , é a estimativa da média populacional da lucratividade de ações do grupo 2. Portanto:

$E(Y_{2t}) = 280,210 + 74,610 = 384,820$ , que é exatamente igual à média da lucratividade do grupo. A interpretação dos demais coeficientes estimados é exatamente equivalente, ou seja, as estimativas dos coeficientes das variáveis apresentadas na tabela 4 são os desvios das médias da variável dependente em cada grupo em relação ao grupo que está na base.

Pelos testes de  $t$  ao nível de significância de 1%, nós aceitamos todas as hipóteses nulas, dos seguintes tipos:

$H_0: \beta_g = 0, (g = 2, \dots, 7)$ , ou seja, em nenhum grupo existe um desvio significativo da média da variável dependente com relação à média do grupo que está na base. Por outro lado o teste de  $F$  testa a seguinte hipótese conjunta:

$$H_0: \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0; G = 7$$

$$H_1: H_0 \text{ não é verdadeira}$$

Esse teste, como foi visto no item 3, é exatamente equivalente ao teste de  $F$  na análise de variância, e deveria apresentar os mesmos resultados de análise feita por Figueiredo, ou seja, a menos de aproximação em casas decimais, o teste de  $F$  da análise de variância realizada por Figueiredo deveria ser igual ao valor de  $F$  calculado pela análise de regressão apresentado na tabela 4. O  $F$  calculado na tabela 4 é 0,591736, ou seja, é menor do que  $F(6,63)$ , com nível de significância de 1%, que é de 3,118; levando à aceitação de  $H_0$ , ou seja, não existe diferenças significativas entre as lucratividades médias dos 7 grupos de ações. Tal conclusão é perfeitamente consistente com aquela de Figueiredo, embora ele tenha utilizado outro método para chegar a ela. Mas, um fato

curioso ocorre; o  $F$  calculado por Figueiredo, a partir da análise de variância, apresenta o valor 1,45, e esse valor é bem diferente do valor por nós encontrado na análise de regressão. Talvez essa discrepância possa ser uma decorrência de erros de cálculo ou de aproximações, embora nos pareça um tanto exagerada.

#### 4.2 Resultados da análise de regressão para o período de baixa

Tabela 5  
Variável dependente: perdas

| Variáveis independentes | Grupo correspondente  | Coefficiente estimado | Teste "t" |
|-------------------------|-----------------------|-----------------------|-----------|
| C                       | Maior VP              | 41,720                |           |
| $X_2$                   | Maior P/L             | 1,500                 | 0,157     |
| $X_3$                   | Menor P/L             | 3,640                 | 0,381     |
| $X_4$                   | Maior negociabilidade | 2,730                 | 0,286     |
| $X_5$                   | Maior VP/PL           | -6,800                | -0,712    |
| $X_6$                   | Maior giro            | 1,340                 | 0,140     |
| $X_7$                   | Seleção aleatória     | 1,940                 | 0,203     |

$$R^2 = 0,0244 \quad F = 0,262998 \quad n = 70$$

Soma dos quadrados dos resíduos = 28 764,8

É interessante verificar que os resultados da tabela 5 são completamente diferentes daqueles obtidos por Figueiredo e levam a conclusões diametralmente opostas. Nesse caso, não é possível apelar para erros de aproximação como forma de explicação dos resultados divergentes. Figueiredo reporta um teste de  $F$  cujo valor é 6,84, o que o leva a rejeitar  $H_0$  e concluir que: em períodos de baixa, aparentemente, a análise técnica permite reduzir a perda com a utilização adequada dos índices técnicos.<sup>7</sup>

Pelo teste de  $F$  da análise de regressão apresentado na tabela 5, nós aceitamos  $H_0$  e concluímos que também em períodos de baixa, a análise técnica não permite reduzir as perdas pela utilização adequada dos indicadores analisados por Figueiredo.

Em resumo, existe uma divergência de resultados que não pode ser explicada apenas por erros de aproximação. Por outro lado, dada a divergência de resultados entre a análise de variância de Figueiredo e da análise de regressão por nós realizada ainda pode pairar uma dúvida com relação à equivalência dos dois enfoques apresentados no

item 3. Para sanar essa dúvida e para clarificar os problemas surgidos nós decidimos refazer as análises de variância reportadas por Figueiredo, com base nos dados por ele apresentados.

Na tabela 6 e 7 são apresentadas as análises de variância correspondentes aos casos: a) período de alta; e b) período de baixa.

Tabela 6  
Análise de variância: lucratividade

| Fonte de variação | Graus de liberdade | Soma de quadrados | Quadrado médio |
|-------------------|--------------------|-------------------|----------------|
| Entre grupos      | 6                  | 1 160 815,9588    | 193 469,3281   |
| Resíduo           | 63                 | 20 597 996,6213   | 326 952,3273   |
| Total             | 69                 | 21 758 812 5901   |                |

$F = 0,5917$

Tabela 7  
Análise de variância: perdas

| Fonte de variação | Graus de liberdade | Soma de quadrados | Quadrado médio |
|-------------------|--------------------|-------------------|----------------|
| Entre grupos      | 6                  | 720,4849          | 120,0808       |
| Resíduo           | 63                 | 28 764,7850       | 456,5839       |
| Total             | 69                 | 29 485,2699       |                |

$F = 0,2630$

Como é fácil perceber os resultados das análises de variância das tabelas 6 e 7 são agora, respectivamente, equivalentes às análises de regressão apresentadas nas tabelas 4 e 5. Senão vejamos: o  $F$  calculado da regressão, apresentado na tabela 4, é 0,591736, enquanto o  $F$  calculado para a análise de variância correspondente, apresentado na tabela 6, é de 0,5917, ou seja, são praticamente iguais. O mesmo acontece com a soma de quadrados dos resíduos, e, portanto, com a variação total (veja tabelas 4 e 6).

Esses resultados também ocorrem para a análise de variância e regressão para o período de baixa da bolsa, os  $F$ 's das tabelas 5 e 7 são iguais, o mesmo acontecendo com a soma de quadrados dos resíduos e, conseqüentemente, com a variação total. Esses resultados são esperados e normais, como foi demonstrado no item 3 deste trabalho. Desta forma, tanto a análise de regressão como a

análise de variância conduzem aos mesmos resultados em termos de testes de hipóteses, o que nos leva a concluir que nenhum dos indicadores técnicos analisados por Figueiredo levam a uma seleção de portfólio com maiores lucros nos períodos de alta e nem a uma diminuição de prejuízos, nos períodos de baixa.

Em suma, o objetivo deste trabalho foi demonstrar a equivalência dos modelos de análise de variância e de regressão. A equivalência teórica foi demonstrada e a aplicação prática dos dois modelos levou a resultados equivalentes em termos de testes de hipóteses. Ao mesmo tempo e, curiosamente, ficou demonstrado que existem erros de cálculo nas duas análises de variância apresentadas por Figueiredo, levando-o a conclusões erradas com relação à eficácia dos indicadores técnicos de bolsa nos períodos de baixa das cotações.

#### Nota do Prof. Orlando Figueiredo

Por desatenção do operador houve erro na alimentação dos dados no computador. Após publicado o artigo e constatado o erro, foi consultada a redação da revista sobre a publicação de uma errata. Em função das modificações necessárias, a redação da RAE aconselhou-me a reescrever o artigo para posterior publicação, programada para o número de maio/junho de 1976.

Como o presente artigo representa uma forma elegante de apresentar uma errata, torna-se desnecessário que o artigo original seja republicado com as devidas correções. De qualquer maneira quero congratular-me com o autor pela forma como o assunto foi tratado em seu artigo. ■

1 Figueiredo, Orlando. Da relevância da análise técnica de ações. *Revista de Administração de Empresas*, Rio de Janeiro, jan./fev. 1974.

2 Os dados utilizados por Figueiredo referem-se à bolsa de São Paulo e foram extraídos da *Revista da Bolsa de Valores de São Paulo*.

3 A formulação das hipóteses  $H_0$  e  $H_1$  é, em termos de notação, diferente da de Figueiredo, ou seja, a sua hipótese nula tem a seguinte formulação:

$H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_G$  ( $g = 1, 2, \dots, G$ ), e  $\mu_g$  é a média de lucratividade do  $g$ -ésimo grupo de ações. A nossa hipótese nula é a mesma de Figueiredo, a diferença é apenas decorrente de diferenças na formulação do modelo de análise de variância. A nossa formulação do modelo é tal que permite uma comparação mais direta com o modelo de análise de regressão. Isso ficará claro no item 3 deste trabalho.

4 Essa formulação permite contornar o problema de indeterminação do sistema de equações normais de mínimos quadrados no cálculo dos valores estimados dos  $\beta$ 's.

5 Veja Johnston, J. *Econometric methods*. McGraw-Hill, 1972. p. 142.

6 Veja Freund, J. E. *Mathematical statistics*. Prentice-Hall, 1962. p. 335 fórmula (14.2.5).

7 Figueiredo, O. op. cit. p. 116.