



## ajuste, por mínimos quadrados, de uma distribuição beta às taxas de fecundidade por idade

Rubens Murillo Marques \*

*Sir, In your otherwise beautiful poem  
(The vision of the Sin) there is a verse which  
reads*

*"Every moment dies a man,  
every moment one is born".*

*Obviously, this cannot be true and I  
suggest that in the next edition you  
have it read*

*"Every moment dies a man,  
every moment 1 1/16 is born".*

*Even this value is slightly in error  
but should be sufficiently accurate for poetry.*

*CHARLES BABBAGE  
(numa carta a Lord Tennyson)*

RESUMO — O trabalho mostra inicialmente a inadequabilidade da utilização do método dos momentos para estimar os parâmetros de uma função Beta ajustada às taxas de fecundidade por idade. É proposta uma solução por mínimos quadrados condicionados de forma que a Taxa de Fecundidade Total seja mantida igual à observada. A solução, interativa via método de Newton Raphson, se mostrou extremamente eficiente convergindo em poucas etapas. Subsidiariamente são obtidas características da distribuição Beta relacionadas ao problema. Os resultados obtidos aplicados a dados do Estado de São Paulo em 1980 mostram, claramente, a obtenção de um melhor ajuste global e em especial na região das altas taxas de fecundidade.

### 1. INTRODUÇÃO

A utilização de modelos estatísticos para a descrição do comportamento das taxas de fecundidade segundo a idade tem sido uma preocupação como bem é salientado por Duchêne e Gillet De Stefano (1974). Estes autores analisam os resultados ad-

---

\* Núcleo de Estudos de População - NEPO/UNICAMP, Campinas, SP e Fundação SEADE, São Paulo, SP.

vindos do ajuste de funções do tipo Hadwiger, Log-normal, Gama, Beta e Polinômio do 3º grau, utilizando o método dos momentos para estimar os parâmetros das distribuições sugeridas e aplicam os resultados a 47 séries empíricas de dados dos mais diversos países e em épocas distintas.

Concluem os autores, entre outras coisas, que a seleção do modelo mais adequado a uma dada série empírica de dados está intimamente relacionada ao valor observado da variância das idades na distribuição das taxas de fecundidade. Assim, a função Gama é a mais indicada quando o desvio padrão da idade na qual as mulheres têm seus filhos é inferior a 6.26; a função polinômio quando este é superior a 7.12 e a função Beta para valores entre estes limites. As demais funções analisadas não apresentam desempenho satisfatório.

O método dos momentos, tão largamente utilizado pela sua simplicidade algébrica, não produz estimadores com propriedades estatísticas desejáveis e leva, como mostraremos, a resultados tendentes a estimar erroneamente importantes variáveis envolvidas no problema.

O presente trabalho propõe um método de obtenção das estimativas de mínimos quadrados para os parâmetros da distribuição Beta (que é a que melhor se ajusta à nossa realidade) e consegue melhorar satisfatoriamente a qualidade do ajuste obtido.

## 2. O MODELO DA DISTRIBUIÇÃO BETA

Sejam  $f_y(y_i)$  as taxas de fecundidade na idade

$$Y = y_i \quad (15 \leq Y \leq 50).$$

Estas taxas são disponíveis idade por idade ou, mais frequentemente, para grupos quinquenais de idade. Em ambos os casos, trabalharemos com os pontos médios das classes correspondentes.

Sejam então:

$$D = \sum_{y_i=15.5}^{49.5} f_y(y_i) \quad (\text{taxa de fecundidade total})$$



$$\bar{y} = \frac{49.5}{\sum_{y_i=15.5}} Y_i f_Y(Y_i) / D \quad (\text{idade m\u00e9dia da maternidade})$$

$$s_Y^2 = \frac{49.5}{\sum_{y_i=15.5}} (Y_i - \bar{y})^2 \cdot f_Y(Y_i) / D \quad (\text{vari\u00e2ncia da idade da maternidade})$$

Fa\u00e7amos a mudan\u00e7a de escala:

$$X = \frac{Y - 15}{35}, \quad 0 \leq X \leq 1$$

Com isso:

$$\bar{x} = \frac{\bar{y} - 15}{35}; \quad s_x^2 = \frac{s_y^2}{1225}$$

Feita esta transforma\u00e7\u00e3o de escala, introduzamos a fun\u00e7\u00e3o:

$$f_{\beta}(x_i) = C \cdot x_i^{p-1} (1-x_i)^{q-1}, \quad p > 0, q > 0, 0 \leq x_i \leq 1, C > 0$$

Esta fun\u00e7\u00e3o \u00e9 mais comumente apresentada sob a forma:

$$f_{\beta}(x_i) = \frac{K}{\beta(p,q)} \cdot x_i^{p-1} (1-x_i)^{q-1}$$

onde

$$\beta(p,q) = \int_0^1 z^{p-1} (1-z)^{q-1} dz \quad \text{\u00e9 chamada fun\u00e7\u00e3o Beta de par\u00e2metro } p \text{ e } q \text{ (} p > 0, q > 0 \text{)}.$$

r\u00e2metro  $p$  e  $q$  ( $p > 0, q > 0$ ).

As fun\u00e7\u00f5es do tipo:

$$f_{\beta}(x|p,q) = f_{\beta}(x) = \frac{K}{\beta(p,q)} x^{p-1} (1-x)^{q-1}, \quad 0 \leq x \leq 1; k, p, q > 0$$

s\u00e3o denominadas distribui\u00e7\u00e3o Beta.

No caso particular de  $K=1$  a distribui\u00e7\u00e3o torna-se a den

sidade de probabilidade Beta,  $\phi_{\beta}(x)$ , haja vista que:

$$\phi_{\beta}(x) \geq 0 \quad \text{e} \quad \int_0^1 \phi_{\beta}(x) dx = 1$$

Esta densidade tem os seguintes dois primeiros momentos (média e variância):

$$E\{X\} = \int_0^1 x \phi_{\beta}(x) dx = \frac{p}{p+q}$$

$$V\{X\} = \int_0^1 (x-E(x))^2 \phi_{\beta}(x) dx = \frac{p}{p+q} \left(1 - \frac{p}{p+q}\right) \cdot \frac{1}{p+q+1}$$

Dependendo dos valores de  $p$  e  $q$  a função densidade Beta apresenta aspectos variados.

Se ambos os parâmetros forem menores do que 1, ela terá a forma de U, se um dos parâmetros for menor do que 1 e o outro maior do que 1 terá a forma de J; se ambos forem iguais a 1, será uma distribuição retangular.

A densidade Beta só será unimodal se  $p > 1$  e  $q > 1$ , e neste caso a sua moda será no ponto:

$$x_M = (p - 1) / (p + q - 2).$$

Os pontos de inflexão (se existirem) estarão dispostos simetricamente em relação à moda, isto é:

$$x_1 = x_M - \Delta \quad ; \quad x_2 = x_M + \Delta$$

onde

$$\Delta = \frac{[(p-1) \cdot (q-1) \cdot (p+q-3)]^{1/2}}{(p+q-2) \cdot (p+q-3)}$$

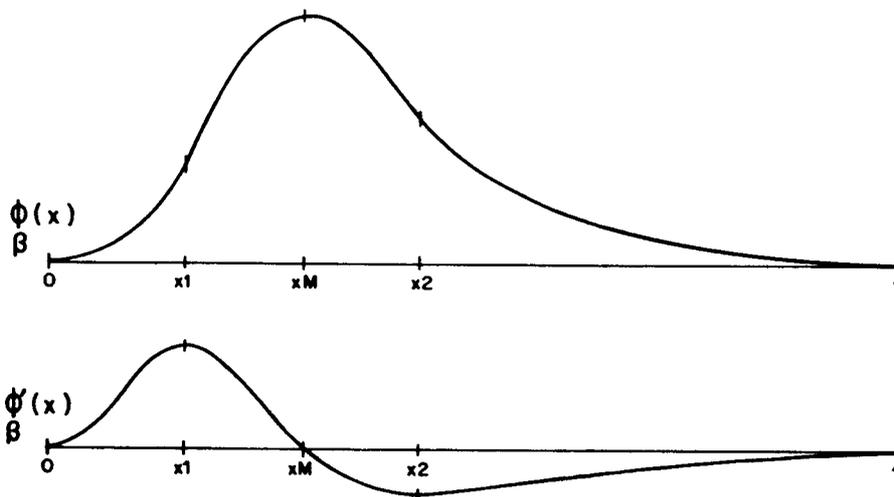
Estes pontos de inflexão são tais que:

$x_1$  não existe se  $p < 2$  ;  $x_2$  não existe se  $q < 2$   
 $x_1=0$  ( $y=15$  anos) se  $p = 2$  ;  $x_2=1$  ( $y=50$  anos) se  $q = 2$   
 $x_1$  é crescente com  $p$  e  $x_2$  é decrescente com  $q$ .



Observando-se o gráfico esquematizado (supondo-se  $p > 2$  e  $q > 2$ ) onde temos a densidade Beta e a sua derivada (velocidade) pode-se notar que a função é crescente com velocidade positiva crescente até  $x_1$  (aonde atinge sua velocidade máxima) continuando crescente, porém, já agora com velocidade positiva decrescente até  $x_M$  (aonde se anula), partindo desse ponto em diante a ter velocidade negativa e decrescente até  $x_2$  (aonde atinge a velocidade mínima) passando daí por diante a continuar com velocidade negativa, porém crescente até se anular em  $x=1$ .

Os pontos de inflexão  $x_1$  e  $x_2$  são, pois, característicos de cada distribuição e nos darão as idades onde ocorrem os maiores aumentos ou quedas nas taxas de fecundidade.



$$\phi'_\beta(x) = \frac{K}{\beta(p,q)} \cdot x^{p-2} (1-x)^{q-2} ((p-1) - (p+q-2)x)$$

A distribuição Beta que ajustaremos às taxas de fecundidade por idade será do tipo:

$$f_{\beta}(x_i) = C \cdot x_i^{p-1} (1-x_i)^{q-1} \quad ; \quad 0 < x < 1, C > 0, p > 1, q > 1$$

diferindo da densidade Beta apenas por um fator multiplicativo e será supostamente unimodal, haja vista as restrições impostas a p e q.

Pode-se verificar facilmente que os momentos desta distribuição não normalizada são os mesmos da densidade Beta.

Evidentemente ela terá também a sua moda e os pontos de inflexão nos mesmos pontos que tinham na função densidade.

O ajuste da função Beta acima depende da estimação dos parâmetros C (ou K), p, q. O valor de C, que é apenas um fator de escala, não apresentará maiores dificuldades. Quanto a p e q podemos utilizar os vários métodos existentes, tais como dos momentos, mínimos quadrados, máxima verossimilhança, etc.

Convém observar que dados p e q, o cálculo da constante  $\beta(p,q)$  pode ser feito pela relação:

$$\beta(p,q) = \frac{\Gamma(p) \Gamma(q)}{\Gamma(p+q)}$$

onde

$$\Gamma(s) = \int_0^{\infty} e^{-z} z^{s-1} dz \quad , \quad s > 0$$

é a função Gama, usualmente também representada na notação de fatorial como  $(s-1)!$ , haja vista ser uma extensão do conceito de fatorial para números reais, positivos e não inteiros.

### 3. OS MÉTODOS DE AJUSTE DA DISTRIBUIÇÃO BETA

#### 3.1 O MÉTODO DOS MOMENTOS

Consiste em igualar os momentos da distribuição a ser ajustada aos momentos observados, obtendo um sistema de equa



ções que resolvido conduz aos estimadores  $p_0$  e  $q_0$  desejados.  
No nosso caso:

$$\int_0^1 f_{\beta}(x) dx = K = D/35$$

$$\bar{x} = \frac{1}{K} \int_0^1 x f_{\beta}(x) dx = p_0/p_0 + q_0$$

$$s_x^2 = \frac{1}{K} \int_0^1 (x-\bar{x})^2 \cdot f_{\beta}(x) dx = \frac{p_0}{p_0 + q_0} \left(1 - \frac{p_0}{p_0 + q_0}\right) \frac{1}{p_0 + q_0 + 1}$$

Este sistema conduz a:

$$p_0 = \frac{\bar{x}^2 \cdot (1 - \bar{x})}{s_x^2} - \bar{x}$$

$$q_0 = \frac{(1-\bar{x})^2 \bar{x}}{s_x^2} - (1-\bar{x})$$

Observe-se que o valor de  $K$  é obtido pelos autores citados impondo-se a igualdade da área sob a curva a ser ajustada à área do histograma dos dados observados, após a transformação da escala.

As principais críticas que realizamos ao método dos momentos como foi utilizado são as de que produzem estimadores sem propriedades estatísticas conhecidas e de que a imposição  $K=D/35$  acarreta que a taxa de fecundidade total ajustada pelo modelo não seja igual à observada, fato este que, em nosso entender, deveria ser preservado.

### 3.2 O MÉTODO DOS MÍNIMOS QUADRADOS

Consiste em obter os valores dos parâmetros que minimizem a soma dos quadrados dos resíduos:

$$R^2(C, p, q) = \sum_{y_i=15.5}^{49.5} \{C \cdot x_i^{p-1} (1-x_i)^{q-1} - f_y(y_i)\}^2, \quad x_i = \frac{y_i - 15}{35}$$

ou simplificando a notação:

$$R^2(C,p,q) = \sum (Cf - f_y)^2$$

Isto implica em resolver o sistema de equações;

$$\frac{\partial R^2}{\partial C} = 0 \quad ; \quad \frac{\partial R^2}{\partial p} = 0 \quad ; \quad \frac{\partial R^2}{\partial q} = 0.$$

No entanto, parece-nos ser mais adequado minimizar a soma de quadrados dos resíduos condicionada a que se obtenha uma taxa de fecundidade total ajustada igual àquela observada, isto é, impor a condição:

$$\sum Cf = D$$

o que implica então em minimizar:

$$R_C^2(p,q) = \sum \left( \frac{D}{\sum f} f - f_y \right)^2$$

isto é, resolver o sistema das equações normais:

$$\frac{\partial R_C^2(p,q)}{\partial p} = 0 \quad ; \quad \frac{\partial R_C^2(p,q)}{\partial q} = 0$$

Adotemos a seguinte notação para simplificar o tratamento algébrico que se segue:

$$f'_p = \frac{\partial f}{\partial p}, \quad f'_q = \frac{\partial f}{\partial q}, \quad f''_{pp} = \frac{\partial^2 f}{\partial p^2}, \quad f''_{qq} = \frac{\partial^2 f}{\partial q^2}, \quad f''_{pq} = f''_{qp} = \frac{\partial^2 f}{\partial p \partial q}$$

$$S = \sum f, \quad S'_p = \sum f'_p, \quad S'_q = \sum f'_q, \quad S''_{pp} = \sum f''_{pp}, \quad S''_{qq} = \sum f''_{qq}, \quad S''_{pq} = S''_{qp} = \sum f''_{pq}$$

Com tal notação temos:

$$\sum \left( \frac{Df}{S} - f_y \right) (f'_p S - f S'_p) = F(p,q) = 0$$

$$\sum \left( \frac{Df}{S} - f_y \right) (f'_q S - f S'_q) = H(p,q) = 0$$

Este sistema de equações não lineares só pode ser resolvido iterativamente.



Convém, no entanto, observar que, caso tivéssemos optado por explicitar a constante  $C$  como  $K/\beta(p,q)$ , a solução de tal sistema se tornaria extremamente mais complexa, senão impossível, pois envolveria a diferenciação da função  $\beta(p,q)$ , o que pode ser visto, implicaria, por sua vez, na obtenção da esperança do logaritmo de uma densidade Beta.

#### 4. O PROCEDIMENTO ITERATIVO DE NEWTON-RAPHSON

Sejam então  $p_n$  e  $q_n$  raízes aproximadas do sistema

$$F(p,q) = 0$$

$$H(p,q) = 0$$

na  $n$ -ésima etapa iterativa ( $n = 0, 1, 2, \dots$ ).

Nestas condições, indicando por:

$$F_n = F(p_n, q_n) \quad , \quad H_n = H(p_n, q_n) \quad , \quad F_{n,p} = \frac{\partial F_n}{\partial p} \quad , \quad F_{n,q} = \frac{\partial F_n}{\partial q}$$

$$H_{n,p} = \frac{\partial H_n}{\partial p} \quad , \quad H_{n,q} = \frac{\partial H_n}{\partial q}$$

e se o Jacobiano:

$$J_n = J_n(p_n, q_n) = \begin{vmatrix} F_{n,p} & F_{n,q} \\ H_{n,p} & H_{n,q} \end{vmatrix} \neq 0$$

teremos as seguintes raízes aproximadas na etapa  $n+1$ :

$$p_{n+1} = p_n - \frac{1}{J_n} \begin{vmatrix} F_n & F_{n,q} \\ H_n & H_{n,q} \end{vmatrix}$$

$$q_{n+1} = q_n - \frac{1}{J_n} \begin{vmatrix} F_{n,p} & F_n \\ H_{n,p} & H_n \end{vmatrix}$$

Os valores iniciais  $p_0$  e  $q_0$  poderão ser as estimativas

fornecidas pelo método dos momentos.

Os elementos ainda necessários para cada etapa n são:

$$F_{n,p} = \Sigma D. \frac{(f'_p \cdot S - fS'_p)^2}{S^2} + \left(\frac{Df}{S} - f_y\right) (f''_p S - fS''_p)$$

$$F_{n,q} = \Sigma D. \frac{(f'_q S - fS'_q)}{S^2} (f'_p S - fS'_p) +$$

$$+ \left(\frac{Df}{S} - f_y\right) (f''_{pq} S + f'_p S'_q - f'_q S'_p - fS''_{pq})$$

$$H_{n,p} = \Sigma D. \frac{(f'_p S - fS'_p)}{S^2} (f'_q S - fS'_q) +$$

$$+ \left(\frac{Df}{S} - f_y\right) (f''_{qp} S + f'_q S'_p - f'_p S'_q - fS''_{qp})$$

$$H_{n,q} = \Sigma D. \frac{(f'_q S - fS'_q)^2}{S^2} + \left(\frac{Df}{S} - f_y\right) (f''_q S - fS''_q)$$

Conquanto este trabalho apresente uma aplicação do método apenas dois casos, testamos a sua utilização para dados de mais de 50 países, com os mais diversos perfis de fecundidade e obtivemos a solução iterativa com um máximo de oito (8) etapas. Assim sendo, cremos que o mesmo possa ser utilizado sem que ocorra alguma dificuldade para a convergência.

## 5. EXEMPLOS

Para ilustrar o método vamos utilizar os dados a seguir:



Tabela 1

Taxas de fecundidade (por mil) segundo classes quinquênis de idade  
Estado de São Paulo, 1980 (\*)

| Idade | Taxa de Fecundidade x 1.000 |
|-------|-----------------------------|
| 15-19 | 73.59                       |
| 20-24 | 187.73                      |
| 25-29 | 188.01                      |
| 30-34 | 129.80                      |
| 35-39 | 71.69                       |
| 40-44 | 25.35                       |
| 45-49 | 4.34                        |

Tabela 2

Taxas de fecundidade (por mil) segundo a idade  
Estado de São Paulo, 1980 (\*)

| Idade | Taxa de Fecundidade<br>x 1.000 | Idade | Taxa de Fecundidade<br>x 1.000 |
|-------|--------------------------------|-------|--------------------------------|
| 15    | 13.94                          | 33    | 113.45                         |
| 16    | 37.45                          | 34    | 100.69                         |
| 17    | 71.52                          | 35    | 88.92                          |
| 18    | 105.79                         | 36    | 82.77                          |
| 19    | 141.89                         | 37    | 70.03                          |
| 20    | 165.65                         | 38    | 58.53                          |
| 21    | 181.84                         | 39    | 54.26                          |
| 22    | 192.11                         | 40    | 38.34                          |
| 23    | 198.73                         | 41    | 32.10                          |
| 24    | 201.58                         | 42    | 23.80                          |
| 25    | 198.81                         | 43    | 17.31                          |
| 26    | 198.13                         | 44    | 12.22                          |
| 27    | 187.50                         | 45    | 8.34                           |
| 28    | 179.47                         | 46    | 5.90                           |
| 29    | 172.67                         | 47    | 2.99                           |
| 30    | 155.12                         | 48    | 2.38                           |
| 31    | 143.28                         | 49    | 1.54                           |
| 32    | 128.75                         |       |                                |

Fontes: Fundação IBGE. Censo Demográfico do Estado de São Paulo, 1980.  
Fundação SEADE. Movimento do registro civil no Estado de São Paulo,  
1980.

Nota: (\*) Os nascimentos foram corrigidos com base no comportamento dos registros atrasados de nascimentos (Giraldelli, B.W. & Wong, L.R., 1984).

Para os dados apresentados obtivemos os seguintes resultados:

Tabela 3

Características dos dados utilizados segundo a amplitude do agrupamento em classes de idade  
Estado de São Paulo, 1980

| Características                    | Classes quinquênnais | Classes unitárias |
|------------------------------------|----------------------|-------------------|
| Média (anos)                       | 27.73                | 27.69             |
| Desvio Padrão (anos)               | 6.63                 | 6.47              |
| Taxa de fecundidade total (filhos) | 3.40                 | 3.39              |

Tabela 4

Características da Função Beta ajustada, segundo o método utilizado e a amplitude do agrupamento em classes de idade  
Estado de São Paulo, 1980

| Características | Classes quinquênnais |                   | Classes unitárias |                   |
|-----------------|----------------------|-------------------|-------------------|-------------------|
|                 | Momentos             | Mínimos Quadrados | Momentos          | Mínimos Quadrados |
| Nº Iterac.      | -                    | 5                 | -                 | 5                 |
| Ganho Real.*    | 0                    | 70.7415           | 0                 | 40.9388           |
| c               | 1.48589              | 2.39718           | 1.86181           | 2.45890           |
| p               | 1.98614              | 2.21109           | 2.09266           | 2.20121           |
| q               | 3.47339              | 3.91628           | 3.67805           | 3.97981           |
| K               | 0.097216             | 0.096197          | 0.096794          | 0.096764          |
| β               | 0.065426             | 0.040129          | 0.051989          | 0.039352          |
| Idade Modal     | 24.98                | 25.27             | 25.14             | 25.06             |
| Pto. Inflexão   | -                    | 16.26             | 15.60             | 16.18             |
| Vel. Máxima     | 41.29                | 33.63             | 36.44             | 35.19             |
| Pto. Inflexão   | 35.05                | 34.28             | 34.68             | 33.94             |
| Vel. Mínima     | -12.16               | -13.84            | -12.95            | -14.09            |
| Fec. Máxima     | 187.94               | 197.19            | 192.39            | 200.35            |
| Idade Média     | 27.73                | 27.63             | 27.69             | 27.46             |
| Desvio Padrão   | 6.63                 | 6.30              | 6.47              | 6.25              |

Nota: \* Ganho relativo = 1 - soma de quadrados dos resíduos pelo Método dos Mínimos Quadrados/soma de quadrados dos resíduos pelo Método dos Momentos.



Tabela 5

Taxas de fecundidade (por mil) observadas e ajustadas pelo Método dos Momentos e dos Mínimos Quadrados, segundo a idade, para os dados da Tabela 1. Estado de São Paulo, 1980

| Idade | Observados |          | Mínimos Quadrados |          | Idade | Observados |          | Mínimos Quadrados |          |
|-------|------------|----------|-------------------|----------|-------|------------|----------|-------------------|----------|
|       | Observados | Momentos | Mínimos Quadrados | Momentos |       | Observados | Momentos | Mínimos Quadrados | Momentos |
| 15.5  | -          | 21.73    | 13.39             | -        | 33.5  | -          | 123.36   | 123.58            | -        |
| 16.5  | -          | 59.69    | 46.50             | -        | 34.5  | -          | 113.31   | 109.76            | -        |
| 17.5  | 73.59      | 91.65    | 79.03             | -        | 35.5  | -          | 99.16    | 96.00             | -        |
| 18.5  | -          | 118.21   | 108.43            | -        | 36.5  | -          | 87.09    | 82.57             | -        |
| 19.5  | -          | 139.85   | 133.81            | 71.69    | 37.5  | 71.69      | 75.29    | 69.71             | -        |
| 20.5  | -          | 156.96   | 154.82            | -        | 38.5  | -          | 63.95    | 57.62             | -        |
| 21.5  | -          | 169.63   | 171.40            | -        | 39.5  | -          | 53.20    | 46.48             | -        |
| 22.5  | 187.73     | 179.14   | 183.67            | -        | 40.5  | -          | 43.21    | 36.44             | -        |
| 23.5  | -          | 184.93   | 191.84            | -        | 41.5  | -          | 34.09    | 27.60             | -        |
| 24.5  | -          | 187.64   | 196.22            | 25.35    | 42.5  | 25.35      | 25.94    | 20.04             | -        |
| 25.5  | -          | 187.60   | 197.11            | -        | 43.5  | -          | 18.86    | 13.78             | -        |
| 26.5  | -          | 185.11   | 194.88            | -        | 44.5  | -          | 12.91    | 8.83              | -        |
| 27.5  | 188.01     | 180.48   | 189.91            | -        | 45.5  | -          | 8.12     | 5.12              | -        |
| 28.5  | -          | 174.00   | 182.58            | -        | 46.5  | -          | 4.50     | 2.56              | -        |
| 29.5  | -          | 165.95   | 173.27            | 4.34     | 47.5  | 4.34       | 2.02     | 1.00              | -        |
| 30.5  | -          | 156.61   | 162.35            | -        | 48.5  | -          | 0.59     | 0.23              | -        |
| 31.5  | -          | 146.24   | 150.20            | -        | 49.5  | -          | 0.04     | 0.01              | -        |
| 32.5  | 129.80     | 135.07   | 137.16            | -        |       |            |          |                   |          |

Tabela 6

Taxas de fecundidade (por mil) observadas e ajustadas pelo Método dos Momentos e dos Mínimos Quadrados, segundo a idade, para os dados da Tabela 1. Estado de São Paulo, 1980

| Idade | Mínimos Quadrados |          |                   | Idade | Mínimos Quadrados |          |                   |
|-------|-------------------|----------|-------------------|-------|-------------------|----------|-------------------|
|       | Observados        | Momentos | Mínimos Quadrados |       | Observados        | Momentos | Mínimos Quadrados |
| 15.5  | 13.94             | 17.26    | 14.31             | 33.5  | 113.45            | 123.81   | 121.61            |
| 16.5  | 37.45             | 53.00    | 49.07             | 34.5  | 100.69            | 110.93   | 107.53            |
| 17.5  | 71.52             | 85.39    | 82.81             | 35.5  | 88.92             | 98.00    | 93.61             |
| 18.5  | 105.79            | 113.43   | 113.03            | 36.5  | 82.77             | 85.25    | 80.11             |
| 19.5  | 141.89            | 136.92   | 138.85            | 37.5  | 70.03             | 72.91    | 67.27             |
| 20.5  | 165.65            | 155.93   | 159.99            | 38.5  | 58.53             | 61.15    | 55.28             |
| 21.5  | 181.84            | 170.64   | 176.44            | 39.5  | 54.26             | 50.16    | 44.32             |
| 22.5  | 192.11            | 181.32   | 188.38            | 40.5  | 38.34             | 40.08    | 34.51             |
| 23.5  | 198.73            | 188.26   | 196.06            | 41.5  | 32.10             | 31.04    | 25.95             |
| 24.5  | 201.58            | 191.78   | 199.82            | 42.5  | 23.80             | 23.11    | 18.68             |
| 25.5  | 198.81            | 192.21   | 200.02            | 43.5  | 17.31             | 16.38    | 12.73             |
| 26.5  | 198.13            | 189.88   | 197.06            | 44.5  | 12.22             | 10.88    | 8.07              |
| 27.5  | 187.50            | 185.12   | 191.35            | 45.5  | 8.34              | 6.59     | 4.62              |
| 28.5  | 179.47            | 178.28   | 183.29            | 46.5  | 5.90              | 3.48     | 2.27              |
| 29.5  | 172.67            | 169.68   | 173.30            | 47.5  | 2.99              | 1.46     | 0.86              |
| 30.5  | 155.12            | 159.63   | 161.76            | 48.5  | 2.38              | 0.39     | 0.20              |
| 31.5  | 143.28            | 148.44   | 149.06            | 49.5  | 1.54              | 0.02     | 0.01              |
| 32.5  | 128.75            | 136.41   | 135.56            |       |                   |          |                   |



Gráfico 1

Taxas de fecundidade (por mil) observadas e ajustadas  
pelo Método dos Momentos e dos Mínimos Quadrados,  
segundo classes quinzenais de idade  
Estado de São Paulo, 1980

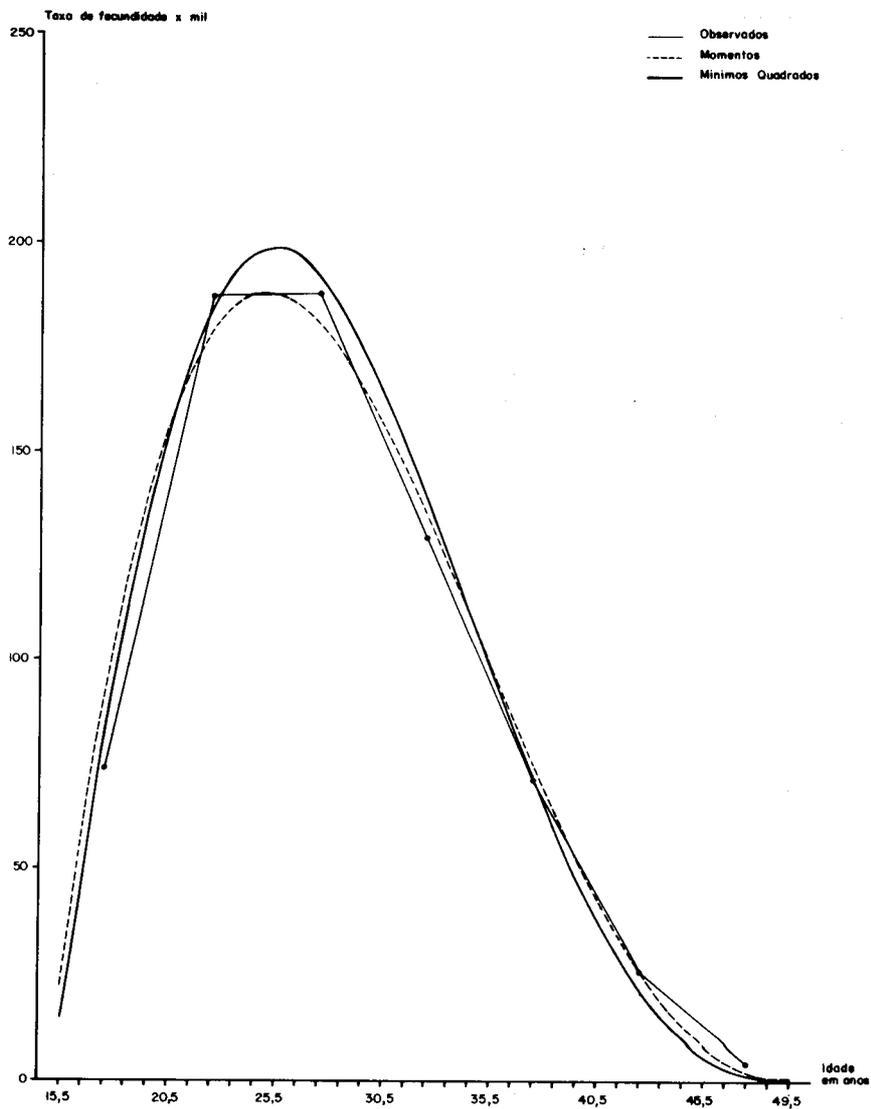
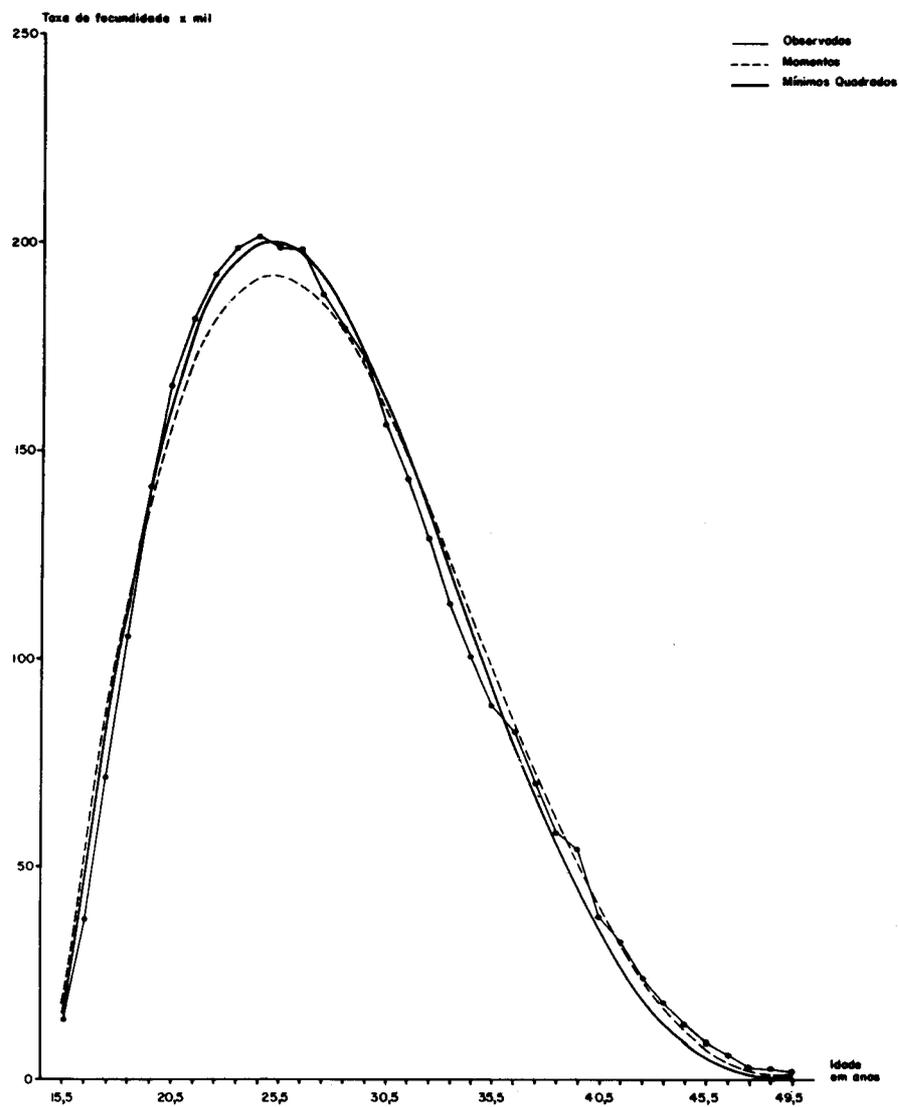


Gráfico 2

Taxas de fecundidade (por mil) observadas e ajustadas pelo Método dos Momentos e dos Mínimos Quadrados, segundo a idade, Estado de São Paulo, 1980





## 6. INTERPRETAÇÃO DOS RESULTADOS

As tabelas e gráficos apresentados mostram nitidamente a inadequabilidade do método dos momentos para a estimação dos parâmetros da função Beta. A solução de mínimos quadrados, além de produzir um ajuste global melhor, como pode ser visto pelo ganho relativo, introduz uma melhoria na região desejável da curva que é a das altas taxas de fecundidade, subestimadas pelo método dos momentos.

Dado o tipo de função utilizada, contendo o produto de duas exponenciais, o ponto de mínimo da soma de quadrados dos resíduos é rapidamente atingido, haja vista a altíssima sensibilidade desta a estes parâmetros.

Devido à correspondência biunívoca entre os pares de valores  $(\bar{x}, s_x^2)$  e  $(p, q)$  podemos determinar, a partir das estimativas de mínimos quadrados obtidos, os valores correspondentes de  $\bar{x}$  e  $s_x^2$ . Isto significa dizer que podemos obter os valores que a média e a variância deveriam ter, a fim de que com estes valores, aplicado o método dos momentos, obtivéssemos a mesma solução.

A Tabela 4 mostra que, independentemente da amplitude do intervalo de classe utilizado, tanto a média quanto o desvio padrão diminuíram. Este enfoque nos permite obter uma interpretação geométrica para a deformação da distribuição inicialmente ajustada pelo método dos momentos em relação à de mínimos quadrados. Ou seja, em ambos os casos a distribuição original foi ligeiramente deslocada para a esquerda e comprimida fazendo com que a sua moda (fecundidade máxima) se elevasse.

Esse fato, aliás, já poderia ter sido parcialmente previsto, pois, como mostra Costa Neto (1980), se os dados realmente provêm de uma distribuição Beta e são agrupados em classes, a variância é sempre superestimada, independentemente do número de classes consideradas (no caso 7 ou 35) desde que  $p > 1.5$  e  $q > 1.5$ . Isto não ocorre necessariamente com a média, dependendo tanto do número de classes quanto dos valores dos parâmetros, sendo, porém, menos sensível do que a variância.

Isto significa que o método dos momentos necessariamente levará a subestimar as taxas de fecundidade nas idades em torno da idade modal, a despeito dos valores estimados da idade modal estarem próximos pelos dois métodos aplicados.

Como a superestimação da variância é uma função decrescente do número de classes, é perfeitamente compreensível o fato de o método dos mínimos quadrados apresentar um ganho maior quando consideramos classes quinquenais de idade.

#### NOTA

Os interessados na obtenção de uma listagem do programa do computador, em linguagem BASIC, para o procedimento apresentado, podem solicitá-lo ao autor, através da Caixa Postal nº 4367, CEP 01000, São Paulo, S.P.

#### REFERÊNCIAS BIBLIOGRÁFICAS

DUCHÊNE, J. & GILLET-de STEFANO, S. 1974. Ajustement analytique des courbes de fécondité générale. Population et Famille, 32:53-93.

COSTA NETO, P.L.O. 1980. "Contribuição ao estudo das correções dos parâmetros de distribuições estatísticas agrupadas em classes". Tese de Doutorado, Escola Politécnica da USP. São Paulo, USP. (mimeo)

GIRALDELLI, Bernadette W. & RODRÍGUEZ WONG, Laura. 1984. O comportamento do registro atrasado de nascimentos no Estado de São Paulo: uma tentativa de correção do sub-registro. Informe Demográfico, nº 13. São Paulo, Fundação SEADE.

ABSTRACT - LEAST SQUARES FITTING OF A BETA DISTRIBUTION TO AGE SPECIFIC FERTILITY RATES - The paper shows at first that using the method of moments to estimate the parameters of a Beta function to fit age specific fertility rates is inadequate. A solution by means of the least squares method is proposed, under the constraint that the implied Total Fertility Rate should equal the observed value. The iterative solution by means of the Newton-Raphson method turns out to be extremely efficient, converging in only a few steps. In addition, some characteristics of the Beta distribution related to the problem are obtained. The results, applied to 1980 data from the State of São Paulo, clearly show a better global fit, especially in the area around the peak of the fertility curve.