

ESCOLA DE PÓS - GRADUAÇÃO EM ECONOMIA  
DO INSTITUTO BRASILEIRO DE ECONOMIA  
DA FUNDAÇÃO GETULIO VARGAS  
*Praia de Botafogo 190*

THE DISTRIBUTION OF EARNINGS IN BRAZIL

Autor: José Julio Senna

SEMINÁRIOS DE PESQUISAS ECONOMICAS I

Data: 20 de fevereiro de 1975

Horário: 13:30h às 15:30h

Sala: 1029

VERSÃO PRELIMINAR

Apenas para circulação interna.

Não pode ser citada sem a pré-  
via autorização do autor.

## I. THEORETICAL FRAMEWORK

### I.1. Schooling and Earnings

In this section we present the schooling model and derive the so-called earnings function, when only schooling is taken into account. This model is the simplest form of human capital models which deal with the problem of earnings distribution. The basic idea behind it is that the length of training in terms of formal education at school is the main source of heterogeneity of labor incomes. Individuals undertake investments in education because they expect that in the future the returns will compensate for the costs of those investments.

In setting up the model we assume that the costs of investments in formal education are restricted to foregone earnings. We also assume that after leaving school the individual enters immediately into the labor market, and does not undertake any further investment in human capital. Moreover, it is assumed that after school the individual receives a constant flow of earnings throughout his working life.

At the time the investment is undertaken, the present value of the flow of future earnings of an individual (after - investment earnings) is only equal to the present value of the flow of earnings in the case of no investment at a certain rate of discount. This rate is the internal rate of discount to such an investment.

In the graph below  $Y_0$  represents the assumed constant level of earnings of an individual who does not invest in education, and  $Y_1$  stands for the level of earnings of an individual who invests one year in education:

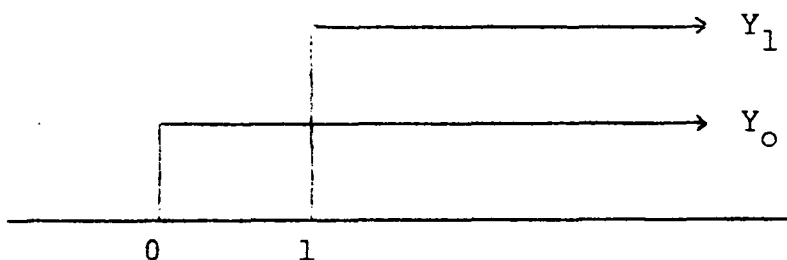


Figure 1

The present value of the flow of earnings of the first individual, evaluated at time 0, is given by:

$$(P V_0)_0 = \frac{Y_0 (1 + r)}{r}$$

and the present value of the second individual, evaluated at time 1, is

$$(P V_1)_1 = \frac{Y_1 (1 + r)}{r}$$

Now, if we discount this value back to period 0, we obtain:

$$(P V_1)_0 = \frac{Y_1 (1 + r)}{r (1 + r)} = \frac{Y_1}{r}$$

where  $(P V_1)_0$  stands for the present value of individual 1 evaluated at time 0.

As mentioned above, to estimate the internal rate of return to the additional year of investment in education we make

$$(P V_0)_0 = (P V_1)_0$$

Therefore,

$$\frac{Y_0 (1 + r)}{r} = \frac{Y_1}{r}$$

which leads to:

$$Y_1 = Y_0 (1 + r)$$

In an analogous way, for two years of investment we have:

$$Y_2 = Y_1 (1 + r) = Y_0 (1 + r)(1 + r) = Y_0 (1 + r)^2$$

and in a general way, for S years:

$$Y_S = Y_0 (1 + r)^S \quad (1)$$

Taking logs, we get:

$$\ln Y_S = \ln Y_0 + S \cdot \ln (1 + r)$$

Since  $\ln (1 + r) \approx r$  for small  $r^{(1)}$ , it follows that: (2)

$$\ln Y_S = \ln Y_0 + r_S S \quad (2)$$

(1) Notice that for values of  $r$  above 0.15 this ceases to be a good approximation, and we may run into specification problems.

(2) Here we add a subscript S to the rate of return, standing for rate of return to schooling, in order to distinguish it from rates of return to on-the-job training, which will appear later on.

Therefore, proportional differences in earnings are directly proportional to absolute differences in time spent at school, and the rate of return is the coefficient of proportionality.

One must now notice that in usual studies of returns to education one can calculate as many rates of return as one wishes (or as many as data availability allows). For example, one can calculate rates of return to each additional year spent at school, at primary, secondary, or university levels. (One can also conceive of rates of return to both private and public education). If one is willing to simplify things, one can estimate rates of return to primary, secondary and university. In the case discussed above, however, what we have is a still simpler / model, in which we are interested in only one rate of return, for schooling in general. In other words, we are dealing with a constant rate of return, which independs from the educational level. For our purposes, this does not seem to be a too drastic assumption.

In addition, in this simplest model, we are assuming that the rate of return does not vary among individuals and that each worker starts out from the same initial level of earnings  $Y_0$ .

Therefore, equation (2) can be rewritten in the following way:

$$\begin{aligned} \ln Y_{S_i} &= \ln Y_{0_i} + r_s S_i && \text{or} \\ \ln Y_i &= a + r_s S_i && (3) \end{aligned}$$

where  $i$  stands for individual.

However, since in reality individuals differ among themselves in terms of rates of return and initial level of earnings, we may impound these differences in a statistical residual, which may contain differentials in school quality as well. One must observe that to the extent that individual differences in rates of return dominate the other types of differences we may have a problem of correlation between the variable  $S$  and

the residual.<sup>(1)</sup>

Therefore,

$$\ln Y_i = a + r_s S_i + u_i \quad (4)$$

If we take the variances in the equation above, we get:

$$\sigma^2 = \text{var}(\ln Y) = \sigma^2(u) + r_s^2 \cdot \sigma^2(S) \quad (5)$$

for constant  $Y_0$  and  $r_s$ . The second part of the right-hand side of equation (5) is the portion of total variance which is explained by the variable formal education. As one knows, the variance of (log of) earnings is an index of inequality of the earnings distribution. From above we observe that the degree of inequality in the distribution is a positive function of the inequality in the schooling distribution and of the magnitude of the rate of return.

Equation (4) generally represents the schooling model. The estimation of such a function will allow us to see to what extent the schooling distribution explains the earnings distribution in Brazil. For estimation purposes we make the hypothesis that  $u$  is independent from  $S$  and  $\ln Y_0$ .

Schooling, however, is just one way of investing in human capital (and hence of acquiring power) and the model can be expanded to include post-school investments. The idea is that individuals continue to improve their skills and earnings capacity after formal education at school, through further investments in themselves. In what follows we take account of this.

(1) Stated more clearly, what we have is:

$\ln Y_i = a + (r + b_i) S_i$ , where  $r$  is the "common" rate of return and  $b_i$  is a component which varies among individuals. Therefore,

$$\ln Y_i = a + r S_i + b_i S_i = a + r S_i + u_i$$

## I.2. Post-School Investments and Earnings

The need to incorporate post-school investments into the framework indicates that the assumption we made earlier regarding the constancy of the flow of earnings after completion of school is not realistic. In fact, we may expect earnings to increase with work experience, as the empirical evidence indicates.

To introduce post-school investments into the analysis, let us define  $Y_{sj}$  as the net earnings of a typical worker in the group with  $s$  years of schooling and  $j$  years of work experience. In the case discussed earlier, in which the individual does not undertake any kind of training besides formal education at school, net earnings will be constant throughout the working period and equal to what we define as potential earnings or gross earnings, or earnings capacity  $E_{sj}$ , neglecting the fact that human capital depreciates over time.

Therefore, in that case we had:

$$Y_{sj} = E_{sj} = Y_0 (1 + r_s)^s$$

Individuals, however, continue to invest in themselves after leaving school, and we may assume that an amount  $C_{sj}$  is invested in period  $j$ . The cost incurred in this investment,  $C_{sj}$ , is either due to foregone earnings, that is, to opportunity costs of time devoted to the investment, on the job or elsewhere, or to direct outlays. In other words, net earnings of an individual in year  $j$ ,  $Y_j$ , is smaller by  $C_j$  than his gross earnings or earnings capacity  $E_j$ , which he could earn if he did not continue to invest in himself through training. The difference  $E_{sj} - Y_{sj}$  is the actual cost of additional investment, and we can write:

$$C_{sj} = E_{sj} - Y_{sj} \quad (6)$$

Hence, when  $C_j$  is only foregone earnings, observed earnings is equal to net earnings ( $Y$ ). The importance of foregone earnings in the case of on-the-job training (compared to investments in formal education) implies that observed earnings is closer to  $Y_j$  than to  $E_j$ . Obviously, the existence of direct costs would lead this procedure to overstate net earnings ( $Y$ ).

We must now notice that the investment  $C_{sj}$  also has a rate of return, which we may call  $r_j$ . If we suppose that the individual invests only for one year (year 0), we have:

$$Y_{S1} = Y_S + r_0 C_0$$

However, if in year 1 he also invests, we have

$$Y_{S1} = Y_S + r_0 C_0 - C_1 = E_{S1} - C_1$$

In more general terms, net earnings in year  $j$  can be written in the following way:

$$Y_{sj} = Y_S + \sum_{t=0}^{j-1} r_t C_t - C_{sj} = E_{sj} - C_{sj} \quad (7)$$

where, from equation (1),  $Y_S = Y_0 (1 + r_s)^S$ .

Now, if we define  $k_{sj}$  as the fraction of potential earnings which is devoted to additional investments, we can write:

$$k_{sj} = \frac{C_{sj}}{E_{sj}} \quad \text{or} \quad C_{sj} = k_{sj} \cdot E_{sj} \quad (8)$$

Consequently,

$$C_{s, j-1} = k_{j-1} \cdot E_{s, j-1}$$

and

$$E_{sj} = E_{s, j-1} + r \cdot C_{s, j-1}$$

$$\therefore E_{sj} = E_{s, j-1} + r \cdot k_{j-1} \cdot E_{s, j-1}$$

$$\therefore E_{sj} = E_{s, j-1} (1 + r \cdot k_{j-1}) \quad (9)$$

$$\therefore E_{sj} = E_{s0} \prod_{t=0}^{j-1} (1 + r_t k_t)$$

Now, if we assume that the rates of return to post-school investment in all periods are the same for any schooling group and equal to  $r_j$ , we have:

$$E_{sj} = Y_0 (1 + r_s)^S \cdot \prod_{t=0}^{j-1} (1 + r_j k_{st}) \quad (10)$$

and using equations ( 6 ) and ( 8 ) above, it comes :

$$Y_{sj} = E_{sj} - C_{sj} = E_{sj} - k_{sj} \cdot E_{sj}$$

$$\therefore Y_{sj} = E_{sj} (1 - k_{sj}) \quad (11)$$

Substituting equation (10) into (11), we have:

$$Y_{sj} = Y_0 (1+r_s)^S \left[ \prod_{t=0}^{j-1} (1 + r_j k_{st}) \right] \cdot (1 - k_{sj})$$

and, taking logarithms:

$$\ln Y_{sj} = \ln Y_0 + \ln (1+r_s) \cdot S + \sum_{t=0}^{j-1} \ln (1+r_j \cdot k_{st}) + \ln (1-k_{sj})$$

and since  $\ln (1 + \lambda) \approx \lambda$ ,

$$\ln Y_{sj} = \ln Y_0 + r_s S + \left[ r_j \sum_{t=0}^{j-1} k_{st} \right] + \ln (1 - k_{sj}) \quad (12)$$

So far we have abstracted from the depreciation phenomenon. This, however, can be incorporated into the model, in the following way.

Let us decompose net investment into gross investment in period  $j-1$ , and that  $\delta$  stands for the rate of depreciation of the human capital stock (hence of earnings  $E_{t-1}^{(1)}$ )

Then, gross earnings in period  $j$  is given by:

$$E_j = E_{j-1} + r C_{j-1}^* - \delta \cdot E_{j-1}$$

---

(1) The rate  $\delta$  affects both the stock of human capital and earnings because a crucial assumption of the model is that the latter is proportional to the level of the former.



$$\therefore \frac{E_j}{E_{j-1}} = 1 + r \cdot \frac{C^*_{j-1}}{E_{j-1}} - \delta$$

$$\therefore \frac{E_j}{E_{j-1}} = 1 + r \cdot k^*_{j-1} - \delta$$

where  $k^*_{j-1}$  may be called the gross investment ratio.

From equation (9) , we observe that:

$$\frac{E_j}{E_{j-1}} = 1 + r \cdot k_{j-1}$$

Therefore,

$$1 + r \cdot k^*_{j-1} - \delta = 1 + r \cdot k_{j-1}$$

and, since the expression above is true for any period,

$$r k_j = r k^*_t - \delta \quad \text{or, more generally,} \quad (13)$$

$$r k_t = r k^*_t - \delta, \quad \text{where } t = (0,1,\dots,j)$$

One must now observe that our procedure in dealing with post-school investments could just as well have been used for investments on formal education <sup>(1)</sup>. Therefore, we can rewrite equation (12) in the following manner:

$$\ln Y_j = \ln Y_0 + \sum_{i=0}^{s-1} r_s \cdot k_i + \sum_{t=0}^{j-1} r_j k_t + \ln (1-k_j) \quad (14)$$

---

(1) One must note that we assumed  $C = E$  during school years, i.e.,  $k_i$  is equal to unity for that period.

Substituting equation (13) into (14) , it comes:

$$\ln Y_j = \ln Y_0 + \sum_{i=0}^{s-1} (r_s k_i^* - \delta) + \sum_{t=0}^{j-1} (r_j k_t^* - \delta) + \ln(1-k_j) \quad (15)$$

Since  $k_i^*$  is approximately equal to unity during school years, that is, net earnings is close to zero during that period, we may write:

$$\ln Y_j = \ln Y_0 + (r_s - \delta) S + \sum_{t=0}^{j-1} (r_j k_t^* - \delta) + \ln(1-k_j) \quad (16)$$

Therefore, if we take depreciation of human capital into account, we need equation (16). If we do not consider depreciation, we need equation (12). Evidently, none of these equations can be used directly for empirical estimations; they require some adaptations, which we shall carry out, for simplicity's sake, for equation (12) only. Later on we shall indicate how depreciation can be introduced.

Let us repeat equation (12):

$$\ln Y_j = \ln Y_0 + r_s S + r_j \sum_{t=0}^{j-1} k_t + \ln(1 - k_j) \quad (12)$$

We must recall at this point that  $k_j$  is the fraction of potential earnings which is invested in each period after leaving school. We may then assume that  $k_j$  declines with  $j$ . The reason for supposing this is that as the individual gets older, the working period which is left over for him to recover the costs of investments in training diminishes, making these investments less attractive. Moreover, the opportunity costs increase as the worker gets more and more experience. Therefore, if we assume that  $k_j$  declines linearly with  $j$ , we obtain:

$$k_j = k_0 - \frac{k_0}{T} j, \quad (17)$$

where  $T$  is the period of positive net investment . (1)

---

(1) Note that when  $j = T$  the net investment ratio is zero, that is, the individual stops investing.

In equation (12) we need, first, to develop the term

$$r_j \sum_{t=0}^{j-1} k_t$$

Applying expression (17) to such a term, we obtain: (2)

$$\begin{aligned} r_j \sum_{t=0}^{j-1} \left( k_0 - \frac{k_0}{T} t \right) &= \\ &= r_j \sum_{t=0}^{j-1} k_0 - r_j \sum_{t=0}^{j-1} \frac{k_0}{T} t = \\ &= r_j \cdot k_0 \cdot j - r_j \cdot \frac{k_0}{T} (0 + 1 + 2 + 3 + \dots + j-1) \end{aligned}$$

In parenthesis we have the sum of a finite arithmetic series, whose value is given by

$$S_n = \frac{(a_1 + a_n)}{2} \cdot n$$

Applying this to the expression above, it comes:

$$\begin{aligned} r_j \cdot k_0 \cdot j - r_j \cdot \frac{k_0}{T} \left( \frac{0 + j - 1}{2} \right) j &= \\ &= r_j \cdot k_0 \cdot j - r_j \frac{k_0}{T} \left( \frac{j^2}{2} - \frac{j}{2} \right) = \\ &= r_j \cdot k_0 \cdot j - \frac{r_j \cdot k_0}{2 T} j^2 + \frac{r_j \cdot k_0}{2 T} \cdot j \end{aligned}$$

---

(2) Recall that  $r_j$  is assumed constant.

Then, equation (12) becomes <sup>(1)</sup>:

$$\ln Y_j = \ln Y_0 + r_s S + r_j \cdot k_0 j + \frac{r_j \cdot k_0}{2T} \cdot j - \frac{r_j \cdot k_0}{2T} + \ln(1-k_j) \quad (18)$$

Let us now make an additional adaptation to the equation above, and use a quadratic approximation in a Taylor expansion for the term  $\ln(1-k_j)$ .

The Taylor expansion is as follows:<sup>(2)</sup>

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots + (-1)^{n-1} \frac{x^n}{n} + \dots$$

which quadratic approximation, in our case, is

$$\ln(1-k_j) = -k_j - \frac{k_j^2}{2} \quad (19)$$

Substituting equation (19) into (18), and adding a residual term we get:

$$\ln Y_j = a_0 + b_1 S + b_2 J + b_3 J^2 + v \quad (20)$$

- ( 1 ) Compare our equation (18) with Mincer's equation (9P), in Mincer 1974b, page 13, or with equation (2b), in Mincer 1974a, page 3-7. In his formulation Mincer disregards the term

$$\frac{r_j \cdot k_0}{2T} j$$

- ( 2 ) See Allen, R.G.D. - Mathematical Analysis for Economists, St. Martin's Press, New York, 1967, page 456.

where :

$$a_0 = \ln Y_0 - k_0 \left( 1 + \frac{k_0}{2} \right)$$

$$b_1 = r_s$$

$$b_2 = r_j k_0 + \frac{r_j k_0}{2T} + \frac{k_0}{T} (1 + k_0)$$

$$b_3 = - \left[ \frac{r_j k_0}{2T} + \frac{k_0^2}{2T^2} \right]$$

In order to incorporate depreciation into the picture, we just repeat the procedure, but instead of starting out from equation (12) we begin from equation (16). The coefficients would then be:

$$a_0 = \ln Y_0 - k^*_0 \left( 1 + \frac{k^*_0}{2} \right)$$

$$b_1 = r_s - \delta$$

$$b_2 = r_j k^*_0 + \frac{r_j k^*_0}{2T} + \frac{k^*_0}{2T^*} + (1+k^*_0) - \delta$$

$$b_3 = - \left[ \frac{r_j k^*_0}{2T^*} + \frac{k^{*2}_0}{2T^{*2}} \right]$$

We must notice here that equation (20) was derived based on the assumption that the investment ratio declines li

nearly with  $j$ . Alternatively, we may now assume a geometric decline in the investment profile, i. e.,

$$k_j = k_0 \cdot e^{-\beta j} \quad (21)$$

when  $\beta$  is the annual percent rate of decline in the investment ratio  $k$ .

Under this hypothesis, we have first to substitute the term

$$r_j \sum_{t=0}^{j-1} k_t$$

in equation (12) by  $r_j$  times equation (21).

Developing that term we obtain

$$\begin{aligned} r_j \sum_{t=0}^{j-1} k_t &= r_j \sum_{t=0}^{j-1} (k_0 \cdot e^{-\beta j}) = \\ &= r_j \cdot k_0 \cdot \sum_{t=0}^{j-1} (e^{-\beta j}) . \end{aligned}$$

We then note that the sum indicated above is the sum of a geometric series whose rate is  $e^{-\beta}$ . Since the sum of such a series can be expressed as

$$S_n = \frac{a_1 - a_n q}{1 - q}$$

where  $a_1$  and  $a_n$  are the first and the last terms, respectively, it follows that:

$$\begin{aligned}
 r_j k_o \sum_{t=0}^{j-1} (e^{-\beta j}) &= r_j k_o \cdot \frac{1 - e^{-\beta(j-1)} e^{-\beta}}{1 - e^{-\beta}} = \\
 &= r_j \cdot k_o \left( \frac{1 - e^{-\beta j}}{1 - e^{-\beta}} \right)
 \end{aligned}$$

Let us now develop the denominator, in series, by MacLaurin's formula:

$$e^{-\beta} = 1 - \frac{\beta}{1!} + \frac{\beta^2}{2!} - \dots$$

Since  $\beta$  is relatively small, we may abandon the quadratic and the subsequent terms:

$$e^{-\beta} = 1 - \beta$$

Then, the denominator becomes:

$$1 - e^{-\beta} = 1 - (1 - \beta) = \beta$$

The term

$$r_j \sum_{t=0}^{j-1} k_t \quad \text{then becomes:}$$

$$r_j \sum_{t=0}^{j-1} k_t = r_j k_o \left( \frac{1 - e^{-\beta j}}{1 - e^{-\beta}} \right) = \frac{r_j k_o}{\beta} (1 - e^{-\beta j})$$

and, substituting this into equation (12), it follows:

$$\ln Y = \ln Y_o + r_s \cdot S + \frac{r_j k_o}{\beta} (1 - e^{-\beta j}) + \ln(1 - k_j) \quad (22)$$

Developing it further, we obtain:

$$\ln Y_j = \ln Y_o + \frac{r_j k_o}{\beta} + r_s \cdot S - \frac{r_j k_o e^{-\beta j}}{\beta} + \ln (1 - k_o \cdot e^{-\beta j})$$

and, making

$$x_j = e^{-\beta j}$$

$$\ln Y_j = \ln Y_o + \frac{r_j k_o}{\beta} + r_s S - \frac{r_j k_o}{\beta} x_j + \ln (1 - k_o x_j) .$$

Applying the quadratic approximation (Taylor series) to

$$\ln (1 - k_o x_j)$$

gives

$$\ln (1 - k_o x_j) = - k_o \cdot x_j - \frac{k_o^2 x_j^2}{2}$$

Hence,

$$\ln Y_j = \ln Y_o + \frac{r_j \cdot k_o}{\beta} + r_s S - \frac{r_j \cdot k_o}{\beta} x_j - k_o \cdot x_j - \frac{k_o^2}{2} \cdot x_j^2$$

Then, another equation to be estimated is a Gompertz function type of equation: ( 1 )

$$\ln Y_j = a_o + a_1 S + a_2 X_j + a_3 X_j^2 + u \quad ( 23 )$$

( 1 )

Note that for estimation purposes we have to impute a value for  $\beta$ , like 0.10 or 0.15.



where ( 1 )

$$x_j = e^{-\beta j}$$

$$a_0 = \ln Y_0 + \frac{r_j \cdot k_0}{\beta}$$

$$b_1 = r_s$$

$$b_2 = - \frac{r_j \cdot k_0}{\beta} - k_0$$

$$b_3 = - \frac{k_0^2}{2}$$

As one can observe not all parameters of the model are determined directly from the estimation of equations ( 20 ) or ( 23 ). Although the estimated average rate of return to schooling is equal to the coefficient of  $S$ , the post-school investment parameters are not directly identifiable. However, in equation ( 20 ), for example, imputing a value for the variable  $T$  (the period of positive net investment), and once we have  $b_2$  and  $b_3$ , we can estimate  $k_0$  and  $r_j$ , since we are left with a system of two equations and two unknowns.

Therefore, the parameters of the equation derived from the model have an economic interpretation. As a matter of fact, this is why it is useful to work within the framework of a theoretical model.

---

( 1 )

Introducing depreciation,  $k_0^*$  replaces  $k_0$  and  $-\delta$  is an additional term in the coefficients  $b_1$  and  $b_2$ .

Additional insights can be obtained if we add the quadratic term  $S^2$  to allow for different rates of return to schooling at different levels of schooling. A negative coefficient for  $S^2$  would suggest that the rate of return to schooling diminishes at higher levels of schooling. In other words, if the coefficient is significant we have an indication that there exists a correlation between  $r_{S_i}$  and  $S_i$  across individuals. The marginal rate of return at each level is then calculated as

$$r_s = \frac{\partial \ln y}{\partial S} = r_1 + 2 r_2 S \quad ( 24 )$$

where  $r_1$  is the coefficient of  $S$ , and  $r_2$  the coefficient of  $S^2$ .

We proceed now to the application of this model to the Brazilian case.

**FUNDAÇÃO GETULIO VARGAS**

**BIBLIOTECA**

ESTE VOLUME DEVE SER DEVOLVIDO À BIBLIOTECA  
NA ÚLTIMA DATA MARCADA

[illegible]

passou p/ estante

N.Cham.      P/EPGE      SPE      S474d

Autor: Senna, Jose Julio de Almeida.

Título: The distribution of earnings in Brazil.



033161

20239

FGV - BMHS

N° Pat. A1336/84

AC 40239

FUNDAÇÃO GETULIO VARGAS  
BIBLIOTECA MÁRIO HENRIQUE SIMONSEN  
AB - 361 84.

JD 33161

**000033161**

