 doi:10.20944/preprints201802.0171.v1

Type of the Paper (Article)

Disabled People in Brazil: Risk Question

Paulo de Oliveira[1,*]

[1] EESC/STT-USP; poliver@usp.br

* Correspondence: poliver@usp.br; Tel.: +55-011-94294-3616

Abstract: Throughout the world, disabilities people have worse health prospects, lower education levels, lower economic participation, and higher poverty rate in comparative terms to people without disabilities. For disabilities people achieve better, more long-lasting prospects, we must empower these people and remove barriers that restrict them from participating in the community have access to quality education, to find decent work and have their voices heard. In statistical terms, a very useful alternative that can provide support and monitoring of public policies in this area is proposing to be used continuously. A risk index called risk index disabled person who is to assess which factors are associated to this risk, as well as the intensity and direction of each of these factors, yielding a final score that can be sorted or classified according to the probability of people acquire a certain disability. In the Brazilian case, we propose the use of techniques such as binary and ordinal logistic regression to select the most significant factors using criteria such as AIC, BIC and DIC. Calculate the risk probability for different disabilities (see, hear, move and intellectual) and number of disabilities to the dataset Sample of respondents for Full Questionnaire at IBGE 2010 Census. In this work, by using stereotype ordinal logistic model with ordinal response, it was possible to improve the fit quality, to be compared with binary response logistic model. By using ordinal response merged the disability risk for different severity degree and amount of deficiencies. Mains conclusions were: i)the model required a smaller number of explanatory variables was intellectual or mental and greater number was disabilities; ii) the most sensitive adjust was using stereotype ordinal logistic; iii) different disabilities aren´t homogenous as the different predictor variables, and finally; iv) higher incidence of risks were noted visual disability, living in the northeast, female, age 80, yellow race, instruction level until incomplete fundamental, works production form their own consumption and high number of children.

Keywords: Disabled people; disabled risk; variables selections; models selections; stereotype ordinal logistic regression

---

1. Introduction

According to the World Health Organization (WHO) in 2010, it is estimated that a little over one billion people worldwide, representing around 15% of world population and in the Brazil case, according to the Brazilian Institute of Geographies and Statistics (IBGE) in 2010. It is estimated that 45.6 million people, representing approximately 23.9% of the population, live with some form of disability. Worldwide, Disabilities people generally tend to have worse health prospects, lower education level, less economic participation and higher poverty rate in comparative terms to people without disabilities.

Disabilities people make up a group excluded that always-aroused different feelings, from revulsion to extreme compassion, and was even considered human or less humanity devoid. Currently, under the social policies inclusion and education, targeted affirmative action became, which seek to assure them their rights in various aspects of life in society [1].

It is believed that low working conditions of disabilities people are due to situations such as difficult access to education, inadequate infrastructure, prejudice, and lack of information and of better accessibility for schools and companies that make these people show a lower education level, which makes them entry into the formal labor market [2].

For disabilities people achieve better and more long-lasting prospects, we must empower these people, remove barriers that restrict them from participating in the community have access to quality education, to find decent work, and have their voices heard [3].

So, you can better assess the needs of disabilities people becomes necessary, best describe this group of people to find answers to questions like how many? Where do you live? How to live? What are the implications of the disability entails access to all these people different human services autonomously and full form? In short, how disability can influence the quality of life of these people?

In statistical terms, shows that there are few formal studies, among which we highlight the data obtained from the census, which allows questions as How do disabilities people are distributed throughout the country? How to evaluate the access of disabilities people in terms of the different services mentioned above? How is the development of disabilities people by comparing them with those who do not have disabilities? What are the variables that contribute most to the cases of disabilities? How are disabilities people compared to those that don´t have disabilities? Response to these and other questions may possibly contribute to better support these people to be better assisted and resources are better managed and optimized by the actions of public policies in this area.

About statistical studies, a very useful alternative that can provide support and monitoring of public policies in this area is proposing to be used continuously. An risk index called *risk index disabled person* who is to assess which factors are more associated to this risk, as well as the intensity and direction of each of these factors, yielding a final score that can be ordered or ranked according to the probability of people acquire a certain disabilities.

The objectives of this work are the implementation of the disability person risk to support and monitor public policies directed to disabled people and their families and the implementation of techniques for selection of variables and models so that it can select among the different variables those that are necessary for the best fit of the model and between the different models determine which one fits best.

In the Brazilian case, we propose the use of techniques such as binary and ordinal logistic regression to select the most significant factors using criteria such as AIC, BIC, and DIC. In addition, calculate the risk probability for different disabilities (see, hear, move and intellectual) to the dataset Sample that consists of 20,800,804 respondents Full Questionnaire at IBGE 2010 Census by state, region and country.

In a previous paper [1] considered as response variables, different disabilities and the existence of at least one disability as binary variables, i.e. whether an individual is a disability person. In this work, we are considering different disabilities incorporating their varying severity degree and number of deficiencies as variables ordinal responses, enabling a better quality in terms of information and tuning the model.

In Section 2 we present a motivation for the problem. We define and characterize the variables to be used describe and define the ordinal logistic stereotype. Variable selection using the Wald test; model selection using AIC, BIC and DIC criteria model and define the disability risk considering the different severity degrees: "*can´t somehow*", "*can, but with greater difficulty*" and "*can with a little difficulty*¨ for visual, hearing and mobility disability, and in the case of intellectual disability suggest the use risk "*having*" or "not have". In section 3 we show the results, discussions and suggestions for future work in section 4 and finally conclusions in section 5.

2. Materials and Methods

2.1. Motivation

So that there is inclusion of disabilities people, it is necessary before to know and what are the factors that most influence the existence of disabilities people. In this paper, we propose the adjustment stereotype ordinal logistic models for the most significant factors using selection criteria such as AIC, BIC and DIC; creation and risk for disability dataset sample who answered the full questionnaire for the 2010 Census of the IBGE.

 doi:10.20944/preprints201802.0171.v1

2.2. Data description

The variables were obtained directly from the questionnaire applied to the sample who answered the Full Questionnaire and can be found on the website www.ibge.gov.br in Census 2010 and micro data with more detail as to your description in [1].

2.3. Logistic Regression

Non-linear mathematical model in which the response variable is qualitative and expressed by two or more levels [4]. Allows to estimate the probability associated to the occurrence of a given event in the face of a set of independent variables, and finally. Aims to estimate the probability of the dependent variable assuming a certain value as a function of those known from other variables. Logistic regression, in this work, is applied in predicting the risk of becoming a person with a disability, classifying these deficiencies as complete, severe or mild.

Among Advantages include:

- Ease of dealing with categorical independent variables;
- Results in terms of probability;
- Classification of individuals into categories;
- A small number of suppositions, and finally;
- High degree of reliability.

2.4. Binary Logistic Regression

To predict whether it has a disability as a function of the related independent variables in topic 2., we apply, in this case, logistic regression, characterized by a binary dependent variable (in this case, whether a person with a disability) and adjust the proposed model in [1].

Binary logistic regression has as advantages: from the mathematical point of view it is extremely flexible, easy to use, allows very rich, direct result interpretation and is able to establish a dependency relationship between a single binary response variable and a set of variables quantitative and qualitative [5].

2.5. Multinomial logistic regression

It is the extension of the logistic regression model developed for binary response variables, for multinomial or polyatomic response variables (three or more categories), it is observed that these categories of the response variable, in this case, need not be ordered and as described in [6]. In this work, we used the multinomial model, considering as a response variable the index of deficiency.

The generalization of the logistic model for response variable with k levels ($k > 2$) is straightforward, allowing its use for discrimination among k classes, but on the other hand, problems can occur in cases of separation in cases of complete separability between classes, which makes possible a unique solution in the likelihood equations and presence of correlation between the independent variables (collinearity).

The logistic regression model, originally developed for binary response variables [1], is extensible for variable politonic responses (three or more categories). For logistic regression with response variable assuming three categories (Y assuming three levels, say 0, 1 or 2), the logistic model will have two logit functions; the ratio of Y = 1 to Y = 0 and the ratio of Y = 2 to Y = 0. In this case, the level Y = 0 was assumed as the basis.

$$g_1(x) = \ln\left[\frac{P(y=1)}{P(y=0)}\right] = \beta_{10} + \beta_{11}x_1 + \cdots + \beta_{1p}x_p$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 February 2018 doi:10.20944/preprints201802.0171.v1

$$g_2(x) = \ln\left[\frac{P(y=2)}{P(y=0)}\right] = \beta_{20} + \beta_{12}x_1 + \cdots + \beta_{2p}x_p$$

From the linear functions $g_i(\mathbf{x})$, whose parameters are estimated by maximum likelihood, it is possible to calculate the conditional probabilities of occurrence of each response variable Y given a vector of observations $\mathbf{x}$, as follows:

$$P(y=0/x) = \frac{1}{1+e^{g_1(x)}+e^{g_{12}(x)}}$$

$$P(y=1/x) = \frac{e^{g_1(x)}}{1+e^{g_1(x)}+e^{g_{12}(x)}}$$

$$P(y=2/x) = \frac{e^{g_{12}(x)}}{1+e^{g_1(x)}+e^{g_{12}(x)}}$$

The generalization of the logistic model for response variables with k levels ($k > 2$) is direct. In the polygonal logistic regression, the probability of a given observation x belonging to one of the classes y_i is estimated directly by the expression:

$$P(y=y_i/x) = \frac{e^{g_1 i(x)}}{1+\sum_{j=1}^{k-1} e^{g_{1j}(x)}}, i=1,2,\cdots,k-1$$

where the logit function, assuming the level y_k as the base, is given by:

$$g_i(x) = \ln\left[\frac{P(Y=y_i/x)}{P(Y=y_k/x)}\right] = \beta_{i0} + \beta_{i1}x_1 + \cdots + \beta_{ip}x_p; i=1',2,\cdots,k-1; g_k(x)=0$$

Considering y_1, y_2, ..., y_k exhaustive categories of variable Y, we can state that $\sum_{i=1}^{k} P(y_i/x) = 1$. Therefore, the probability of an observation x belongs to class y_k, defined by $P(y_k/x)$, can be obtained by the difference: $P(y_k/x) = 1 - \sum_{i=1}^{k-1} P(y_i/x)$.

The logistic model needs to estimate *k-1* parameter vectors $\beta_i' = \lfloor \beta_1, \beta_2, \cdots, \beta_p \rfloor$, corresponding to *k-1* categories of variable **Y**. The *k-th* category is assumed as the base. The process of estimating logistic regression parameters is based on the maximization of the likelihood function $\ell(x.\beta)$.

Under the hypothesis of the sample being representative of the study population, a model is obtained that maximizes the chances of classifying all observations of the population in the classes y_i, which really belong. Since the equations derived from the likelihood function are nonlinear, there is a need to use numerical methods to find a solution. These procedures are iterative, and for this study we use the Multinomial Logistic Regression procedure.

2.6. Ordinal logistic regression

Many of the variables of study in the humanities and social sciences are ordinal. Often, the dependent variable takes discrete values, or sortable categories but whose distance between them is not known, nor constant. For example, in epidemiological and severity degree of disability studies to see, hear or move as established in the sample questionnaire in Census 2010 IBGE which can be classified into "*can´t somehow*", "*can, but with greater difficulty*", "*can, but with some difficulty*", and, finally, "*presents no problem*" to hear, see or move. In the case of intellectual disability was divided into "*have*" or "*haven´t*".

Among the possible models for ordinal logistic regression is possible to highlight:

- The proportional odds which is more appropriate and easier interpretation model when the response variable to be considered is a continuous variable that was categorized, continuation ratio model that is used in cases there is specific interest in a category of response variable;

- Partial proportional odds model that allows some covariates can be modeled with the assumption of proportional chance, and for the other variables in that this assumption is not satisfied, are included in the model specific parameters ranging compared for the various categories, and it is an extension of the proportional odds model, and, lastly,
- Stereotype model proposed by [7] that is used in situations where the response variable is an ordinal variable it isn´t a discrete version of a continuous variable.

For this work, we have as the response variable: disabilities, visual, hearing, intellectual and move that these are ordinal variables that aren´t continuous variables version, in view of this, we adopt in this work the stereotype ordinal logistic regression model.

2.6.1. Specification of the stereotype model

Suppose that the dependent variable is constituted by J categories (m = 1, ..., J) and K consider regressors (j = 1, ..., K). The model defines stereotype at an early stage with the multinomial regression model to which is added the condition $\beta_{m|J} \equiv \phi_m \hat{\beta}$ where J is the reference category, that is, have the regression model is given by the multinomial.

$$\mathrm{Prob}(\mathrm{y} = \mathrm{m}|\mathrm{x}) = \frac{\exp(\beta'_{m|J} x)}{\sum_{j=1}^{J} \exp(\beta'_{m|J} x)}, \text{ With m = 1, ..., J.} \tag{1}$$

Replacing $\beta_{m|J} = \phi_m \tilde{\beta}$ in equation (1) results in stereotype model that can be written as:

$$\mathrm{Prob}(\mathrm{y} = \mathrm{m}|\mathrm{x}) = \frac{\exp(\phi_m \tilde{\beta}' x)}{\sum_{j=1}^{J} \exp(\phi_1 \tilde{\beta}' x)} = \frac{\exp(\phi_m \tilde{\beta}_0 + \phi_m \tilde{\beta}_1 x_1 + \ldots + \phi_m \tilde{\beta}_k x_k)}{\sum_{j=1}^{J} \exp(\phi_m \tilde{\beta}_0 + \phi_m \tilde{\beta}_1 x_1 + \ldots + \phi_m \tilde{\beta}_k x_k)}. \text{ With } m = 1, \ldots, J. \tag{2}$$

For some parameters of equation (2) that aren´t identifiable, we consider restrictions $\phi_m \tilde{\beta}_0 \equiv \theta_m (m = 1, \ldots, J)$, where $\phi_J = 0$; and $\phi_m \tilde{\beta}_j \equiv -\theta_m \beta_j (m = 1, \ldots, J$ and $j = 1, \ldots, k)$, where $\phi_J = 0$ and $\phi = 1$.

Thus, from equation (2), the stereotype model can be written as follows:

$$\mathrm{Prob}(\mathrm{y} = \mathrm{m}|\mathrm{x}) = \frac{\exp(\theta_m - \phi_m \beta' x)}{\sum_{j=1}^{J} \exp(\phi_1 \beta' x)}, \tag{3}$$

with m = 1, ..., J and where $\theta_J = 0$, $\phi_J = 0$ e $\phi = 1$.

2.6.2. Interpretation of the estimated coefficients

Applying logarithm function (3) for any two categories, we have:

$$\log\left[\frac{p(Y = q/x)}{p(Y = r/x)}\right] = (\theta_q - \theta_r) - (\phi_q - \phi_r)\beta' x. \tag{4}$$

Applying the exponential function to the equation (4) below:

$$\Omega_{q/r} = \frac{p(Y = q/x)}{p(Y = r/x)} = \exp\{(\theta_q - \theta_r) - (\phi_q - \phi_r)\beta' x\}. \tag{5}$$

Equation (5) allows us to assess the odds ratio before and after you add a drive to the variable x_j, i.e.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 February 2018 doi:10.20944/preprints201802.0171.v1

$$\frac{\Omega_{q/r}(x, x_k + 1)}{\Omega_{q/r}(x, x_k)} = \exp\{(\phi_r - \phi_q)\beta_x\} \quad (6)$$

The value obtained in expression (6) can be interpreted as adding a unit in variable X_k, the odds ratio varies from r categories $\exp\{(\phi_r - \phi_q)\beta_k\}$, holding all other variables constant.

2.6.3. *Estimation of the estimated coefficients*

The model parameters are estimated by stereotype Maximum Likelihood method, in which the estimators are obtained by the equations given in (7) the following system:

$$p_i = \begin{cases} \mathrm{Prob}(y_i = 1 | x_i, \phi, \theta) & \text{se } y_i = 1 \\ \vdots & \\ \mathrm{Prob}(y_i = m | x_i, \phi, \theta) & \text{se } y_i = m \\ \vdots & \\ \mathrm{Prob}(y_i = J | x_i, \phi, \theta) & \text{se } y_i = J \end{cases} \quad (7)$$

where, p_i is the probability of observing any value of y, and the was defined in expression (3). When we assume that the sample is independent and identically distributed, the likelihood function is given by (8) below:

$$\mathrm{L}(\beta, \phi, \theta | y, x) = \prod_{i=1}^{N} p_i = \prod_{m=1}^{J} \prod_{y=m} \mathrm{Prob}(y = m | x, \phi, \theta) \quad (8)$$

where, $\prod_{y=j}$ indicates that the multiplications on all cases where $y = m$ (m = 1, ..., J). Applying logarithm of the likelihood function obtained in (8) we obtain the logarithm of the likelihood function given in (9) below:

$$\log(\mathrm{L}(\beta, \phi, \theta | y, x)) = \sum_{m=1}^{J} \sum_{y=m} \log[\mathrm{Prob}(y = m | x, \phi, \theta)] \quad (9)$$

The parameters ϕ´s and θ´s of equation (9) are estimated by the Newton-Raphson method.

The odds ratio will tend to increase, since the weights can be constructed by ranking. Thus, the effect of covariates on the first odds ratio is smaller than the second effect, and so on.

The determination of these weights can be made a priori, being estimated for a pilot study or set of appropriately chosen values.

For example, the number of deficiencies, which has an individual and can vary from 0 to 4 deficiencies in this case would be five possible response.

As the fit quality for the ordinal models, this can be checked using the Pearson or deviance tests. These tests involve the creation of a contingency table in which the rows consist of all possible configurations of the covariates in the model and the columns are the ordinal response categories [8]. The expected counts (E_{ij}) in this table are express by $E_{lj} = \sum_{l=1}^{N_L} \hat{p}_{lj}$, where N_L is the number of individuals classified on line l and $\hat{p}_{lj}$ is the probability that an individual in line lj have the response calculated from the model adopted.

Pearson's test to assess the adequacy of the fit compares these with the expected counts observed by the formula

$$\chi^2 = \sum_{l=1}^{L} \sum_{j=1}^{k} \frac{(O_{lj} - E_{lj})^2}{E_{lj}} \quad (10)$$

The deviance statistic also compares observed counts (o_{lj}) and expected, but using the formula:

$$D^2 = 2\sum_{l=1}^{L}\sum_{j=1}^{k} O_{lj} \log \frac{O_{lj}}{E_{lj}} \quad (11)$$

Tests to evaluate the fit of the model are based on approximations of the statistics (10) and (11) for the chi-square distribution with *(G - 1) (k - 1) p* freedom degrees, where *L* and *k* are as defined above and *p* is the number of covariates in the model. Significant differences lead to the conclusion of fit lack of the model to the data studied.

$$W = (\hat{p} - \hat{p}_0)' \hat{V}_p^{-1} (\hat{p} - \hat{p}_0) \quad (12)$$

where, $\hat{V}_p$ is the consistent estimator of the variance-covariance matrix estimator $\hat{p}$ of the vector of proportion $\hat{p}$. An estimator $\hat{V}_p$ can be obtained by linearization.

2.6.4. Significance test for the model

The Wald test for the parameters individually, can be obtained by comparing the maximum likelihood estimate of a given coefficient to its estimated standard error (based on the asymptotic distribution of maximum likelihood estimators). Thus, the null and alternative hypotheses of the test are respectively:

$$H_0 : \hat{\beta}_j = \beta_j^* \quad \text{vs} \quad H_1 : \hat{\beta}_j \neq \beta_j^* \qquad (\text{j=2,…,k}),$$

being the respective statistical under the null hypothesis:

$$T = \frac{\hat{\beta}_j - \beta_j^*}{\sqrt{\text{var}(\hat{\beta}_j)}} \sim \text{N}(0,1)$$

parameter is relevant to explain the behavior of the dependent variable. In rejecting H_0 for a α, level significance we conclude that the estimated parameter is statistically different β_j^*. Generally, if it uses that $\beta_j^* = 0$, given these conditions, we conclude that the parameter is relevant to explain the behavior of the dependent variable.

2.6.5. Variables´ Selection

Select means to choose a subset of variables that retains the most important predictor variables excluding others, such as seeking to avoid multicollinearity problem and that this subset adjustment as well as the model with all variables and containing the most important predictor variables.

Among the different strategies that can be used to select variables include forward stepwise and backward stepwise. A forward stepwise regression starts with the constant β_0, and, sequentially adds to the model the predictor X_i most correlated with *Y* so that improves the fit according to the evaluation of the F statistic; and the introduction of variables for when you cannot produce a greater *F* statistic that 90 or 95 percentile distribution $F_{1,N-k-2}$, where *N* is the sample size and *k* is the number of variables. Already, the backward stepwise selection strategy starts with the model with all independent variables and sequentially goes excluding variables using the *F* statistic to choose the predictors to eliminate.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 February 2018 doi:10.20944/preprints201802.0171.v1

The predictor that has the lowest F statistic is eliminated and the process for each predictor eliminated when the model has an F value greater than the 90 percentile or 95 distribution $F_{1,N-k-2}$, in the case of this paper we have used the forward backward and Wald statistics.

In stereotype ordinal logistic regression, significance is ensured through the TRV (likelihoods ratio test). Thus, at each step of the procedure the most important variable in statistical terms, is that which produces the greatest change in log likelihood relative to a model that doesn´t contain the variable [9].

After the parameters have been estimated, the next step is to assess whether the covariates used and available for modeling are statistically significant with the event modeled, such as the condition of an individual being a disability person.

One way to test the significance of the coefficient of a given covariate is to compare the observed values of the response variable with the predicted values obtained by models with and without the variable of interest [10].

The comparison between the observed and predicted values is done from the likelihood ratio, which is broadly applicable to the estimation of maximum likelihood test.

For test $H_0 : \theta \in \Theta_0$ *versus* $H_0 : \theta \in \Theta_0^c$. We calculate the Statistics [11]: $\lambda(x) = \frac{\sup_{\theta_0} L(\theta/x)}{\sup_{\theta} L(\theta/x)}$. For $n \to \infty, -2 \ln \lambda(x) \to \chi_v^2$.

Where, v is obtained from the difference between the number of parameters in the model tested and the amount of the saturated model parameters [12].

The quality of the fitted model can be verified by comparing the observed and predicted values for the response variable (in this case may be one of various deficiencies already mentioned).

In choosing a model, on the one hand, we should try to include as many independent variables as possible to improve prediction, on the other; we want to include a minimum number of variables for cost problems and simplicity [8].

According to [13] defines select the best model as the commitment to reconcile these two goals (to incorporate a certain number of variables that can improve the predictability of the model, while discarding variables that aren´t significant to simplify the model and reduce costs). This selection involves a dose of subjectivity; the result may be different if you change the procedure for selection.

2.6.6. Selection of models

Select a model means, after the formulation and setting of different plausible models, select the model that ¨best fits¨ the data of a certain experiment according to a given criterion adopted [14].

In Statistics, there is a vast literature on model selection [15, 16, 17]. An alternative to model selection is to use methods based on the likelihood function that provides various statistical measures that help in comparing different models.

The most common of these measures are the following:

- AIC, proposed by [18] and [19] with penalty given after deducting the value of twice the difference between the number of parameters between the two models;
- BIC discussed by [18] and with the penalty value of twice the number of parameters between the two models multiplied by the logarithm of the sample size, and finally;
- DIC also discussed by [18], and the penalty is given by the sum of the value of the difference between the number of parameters between the two models [6].

In this work, we use the AIC, BIC and DIC criteria for select the model with the lowest value for each of these criteria.

2.7. Epidemiology

According to the IEA (International Epidemiology Association) defines epidemiology as the study of factors that determine the frequency and distribution of diseases in human collectivities.

According to [20], epidemiology is essentially a population science, which relies on the social sciences for understanding social structure and dynamics, in mathematics for statistical notions of

 doi:10.20944/preprints201802.0171.v1

probability, inference and estimation, and in the biological sciences for substrate knowledge where the observed manifestations will find individual expression.

Motivated by the increasing complexity and considering the breadth of its current practice, a single and precise definition of epidemiology as a scientific field is not possible. In a simplified way, we can conceptualize it as:

Science that studies the health-disease process in society;

- Analyzing population distribution and determinants of risk, diseases, diseases and health-related events;
- Proposing specific measures to prevent;
- Control or eradicate diseases;
- Damages or health and protection problems;
- Promotion or recovery of individual and collective health:
- Producing information and knowledge to support decision making in the planning, and finally;
- Administration and evaluation of health systems, programs, services and actions [21].

Epidemiology is a basic public health discipline focused on the understanding of the health-disease process within populations, an aspect that unlike the clinic, which aims to study this same process, but in individual terms.

As a science, epidemiology is based on causal reasoning; already as a public health discipline, is concerned with the development of strategies for actions aimed at protecting and promoting community health.

Epidemiology is also an instrument for the development of policies in the health sector. Its application in this case must consider the available knowledge, adapting it to the local realities.

Among the main uses of epidemiology, it is possible to mention:

- Analyzing the health situation; identify profiles and risk factors; undertake the epidemiological evaluation of services;
- Understand the causality of health problems; describe the clinical spectrum of diseases and their natural history;
- Assess how health services respond to the problems and needs of populations; to test the effectiveness, effectiveness and impact of intervention strategies, as well as the quality, access and availability of health services to control, prevent and treat health problems in the community;
- To identify risk factors for a disease and groups of individuals that are at greater risk of being affected by a disease;
- Define the modes of transmission to identify and explain patterns of geographical distribution of diseases;
- Establish methods and strategies for controlling health problems;
- Establish preventive measures; assist the planning and development of health services, and finally;
- Establish criteria for health surveillance.

For the issue of deficiencies, we are considering in their studies the epidemiological visions under the social aspects, under aspects of accessibility, assistive technology among others, and; from the point of view of prevention, treatment and control.

2.8. Risk

Risk is a recent and essentially modern term. It reflects the reorientation of people's relationships with future events. If before the modern age the danger implied fatality, now it comes to be signified as a future event.

The word risk dates from the fourteenth century, gaining the connotation of danger only in the sixteenth century. Among the various meanings of risk, [22] highlights two dimensions. The first refers to what is possible or probable to understand the regularity of phenomena. The second is in the sphere of values and presupposes the possibility of loss of something precious.

The incorporation of the notion of risk was the result of social and technological changes. It is articulated to the laicization of society and to the transformations in the economic relations of commercial capitalism, to open trade and to the development of unprecedented policy structures, such as sovereignty over national territories. It is in this context that also emerges the theory of probability, another phenomenon associated with the notion of risk. Probabilistic thinking favored

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 February 2018 doi:10.20944/preprints201802.0171.v1

the necessary terrain to think of risks as possible management [22]. The risk calculation is closely related to the conformation and valorization of security

In terms of health, the risk is individualized in what is called self-management (it is assumed that people, using sufficient information, adapt their behaviors, eliminating all risks and thus achieving full health [23].

According to [24], the word risk has been gaining frequency in medical journals in the last three decades and suggests that this increase may be related to many factors, including the development of disciplines for risk calculations, expressed as statistical probabilities; the recent development in computational technology; risk management; safety and health promotion. But what is most important in this process is that control of the hazard, previously related to unpredictable, fatalistic factors, now appears as amenable to human control.

2.8.1. Epidemiological risk.

In health, some studies on risk concentrate the focus in epidemiology. Synthetically, epidemiological risk can be defined as the probability of occurrence of a certain health-related event, estimated from the event that occurred in the recent past. Thus, the risk is calculated by quantifying the number of times the event occurred divided by the potential number of events that could have happened. Thus, the risk of becoming a disabled person in each population or group of people is the number of people with disabilities in the previous period by the number of people in that period, since any person or all could potentially person with a disability.

The constitution of the concept of epidemiological risk and the method incorporated by medical research end up defining lifestyles, producing meanings that guide behavior; [23]. In this way, it is possible to observe the presence of the individual in the form of a self-control.

In this work, we are considering the risk of a person becoming a person with disabilities, encompassing a set of health and social factors.

2.8.2 risk disabilities

According to the WHO:

- The prevalence of disabilities people is high;
- The number of disabilities people increases due to the aging population and due to the overall improvement of the chronic health conditions associated with disability such as diabetes, cardiovascular disease and mental illness;
- Several experiments in which the deficiency resulting from the interaction between health conditions, personal and environmental factors ranging widely, and, finally;
- Vulnerable populations whose prevalence varies with the conditions of each country, purchasing power, working conditions and training level. Factors such as these are considered as risks for people to be disabled, which in turn can aggravate the situations mentioned above.

In this scenario, has emerged reasons justifying the need to assess the well-being or life quality of disabled people, we propose the creation of the index risk disabilities person made up by weighting the responses of different variables obtained from micro data of the Census IBGE and selected as significant after applying the backward stepwise methodology in setting the stereotype ordinal logistic type regression for each studied disability.

This methodology has emerged gradually from simpler to more complex techniques such as multivariate techniques such as factor analysis.

2.9. Variables description

This project will be considered the following variables allocated in the following topics:

- identification: identification number, state, region, gender, age, race, birth registration, nationality, was born in this township, was born in this federal state or country foreign birth, duration of residence at FS (Federation State), time of residence in the municipality;
- disability: permanent difficulty seeing, hearing, and walk; intellectual;
- education: to read and write, attends school or daycare, course that nowadays attending, other graduation, course who attended and highest education level;
- family: live in the company of the spouse or partner(a), the respondent's relationship status, nature of marriage, marital status and number of children;

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 February 2018 doi:10.20944/preprints201802.0171.v1

- housing conditions: occupation situation, type of property, rental value, permanent material of outer walls, number of rooms, resident/room density, number of bedrooms, occupant density/dorm, number of bathrooms, toilets, sewage, water supply form, destination trash and electricity;
- other existing goods in residence: radio, television, washing machine, fridge, mobile phone, landline phone, PC, internet access, motorcycle and car, and finally;
- work: how work had, this work was how many people employed, contributed to the government pension, income, returns home from work daily, travel time between home and work situation of occupation, position occupancy in main and secondary employment position.

For this study we are also proposing the creation of the following variables with their respective categorizations: the topic work: income in minimum wages categorized (mw, which at the time of completion of the 2010 Census was 510 reals) to 1, if win between 0 to 1 mw; 2, of 1 to 3 mw; 3, of 3 to 7 mw; 4, of 7 to 15 mw; and, finally; 5, of 15 or more mw.

The topic ID: ID (number questionnaire or individual) and categorized age (1, if between 0 to 15 years; 2, if have between 15 to 65; and, finally; 3, if have 65 years or more.

In the family topic: number of children categorized (1 for childless; 2 for number of children between 1 and 2; 3, for the number of children between 3-5; and, finally, 4, number of children from 6 or more).

By last, on the topic disabilities: deficiencies (number of disabilities that each individual possesses, and ranges from 0 to 4); defic1 (0, if you do not have disabilities and 1, if it has at least one disability); rental categorized minimum wages (1, for values between 0 rents and mw 0.5; 2, between 0.5 and 1 wm; 3, between 1 and 2 mw; 4, between 2 and 4 wm; and, finally; 5, to values larger than 4 mw); number of categorized rooms (1, if the number is 1 to 3 rooms; 2, if four rooms; 3, if 5 rooms; 4, if 6; 5, if seven rooms; and finally; 6, if the house has 8 rooms or more); number of bedrooms categorized (1, if the house has one bedroom; 2, if the house has 2 bedroom; 3, if the residence has 3 bedrooms; 4, if the residence has 4 bedrooms or more); categorized by room occupant density (1, if the density is between 0 and 1 occupant per room; 2, if it is between 1 and 2 residents per room; and finally, 3; if the home has a density greater than 2 residents per room), and, categorized by resident per bedroom (1, if the residence is between 0 and 1 residents per bedroom; 2, if have between 1 and 2; 3, if have between 2 and 3; 4, if have between 3 and 4; 5, if have between 4 and 5; and, finally; 6, if has 5 or more).

3. Results

For this study, we used stereotype ordinal logistic regression analysis for each of the following response variables logistic regression:

- Deficiencies, which represents the number of disabilities that each person possesses and can take a value among 0-4 deficiencies;
- Deficiencies to see, hear and move considering the categories: 0, "*for those who can´t somehow*"; 1, "*for those who can, but with greater difficulty*" 2, "*for those who can, but with some difficulty*" and by end 3, for whom "*presents no problem*", and finally;
- Intellectual Disability considering the categories "*have*" or "*haven't*".

For each of the following sections: identification, education, family and work; the model consisting of all significant variables and for each block in adjusted models were applied:

a) Methods of backward stepwise for variable selection and the variables that shape not considered significant by the Wald test in each step were excluded;

b) Repeat the analysis until no more variables to be excluded;

c) For each of these adjustments models was calculated selection criteria AIC, BIC and DIC;

d) Selection of the best model among the different final models for each of the different number of shortcomings and deficiencies for each of the following criteria: AIC, BIC and DIC, and finally;

e) The risk of everyone having a disability for different severity degrees and number of disabilities was calculated.

In Figures 1-8 show in the items:

a) Graphically the risk of being with one disability (represented by p_1 in blue dots), two (represented by red dots in p_2), three (represented by green dots on p_3), four deficiencies (represented by purple dots on p_4) and at least one disability (represented by black dots in p_t), and;

b) to be visual disability person for each of the different severity degree: "*can´t see any way*" (represented by blue dots on p_3); "*can, but with greater difficulty*" (represented by red dots in p_2); "*can,* *but with a little difficulty*" (represented by p_1 in green dots), and finally, "*the be a person with a visual disability*" (represented by purple p_t in points).

For the variables region in Figure 1, Figure 2 in sex, age in Figure 3, Figure 4 race, education level in Figure 5, the main work in Figure 6, in Figure 7 categorized income and number of children in Figure 8.

For Figure 1, the regions considered were: 1 - "*north*", 2 - "*northeast*", 3 - "*southeast*", 4 - "*South*" and 5 - "*Midwest*".

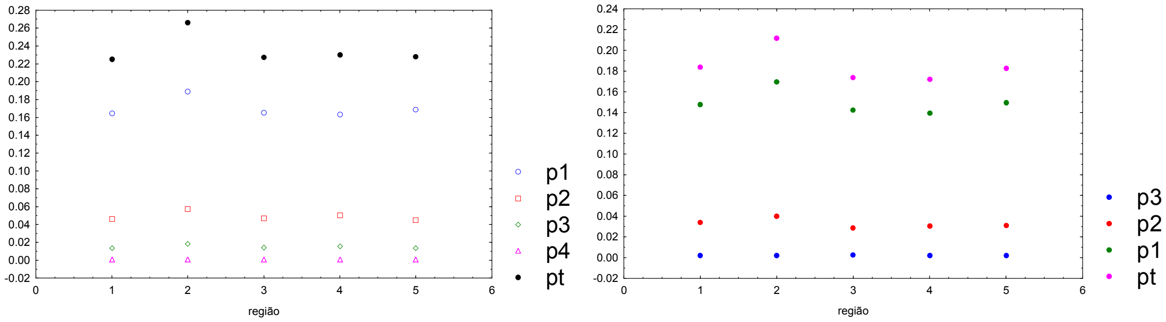


Figure 1. Graphics of probability of occurrence: a) a number of deficiencies, b) visual disability according to their severity degree for the region variable.

In the graphics of Figure 1 we note that the highest incidence risks in item a) of disability and item b) of visual disability lies in the northeastern region to all different amount of disability and all different severity degrees. While the lower incidence rates in the number of deficiencies lies in the west central region and b) the lower incidence of risk of visual disability is in the southern region.

Figure 2 present in a) risk of disabled person and b) the incidence risk of visual disability person considering sexes 1 - male and 2 - female.

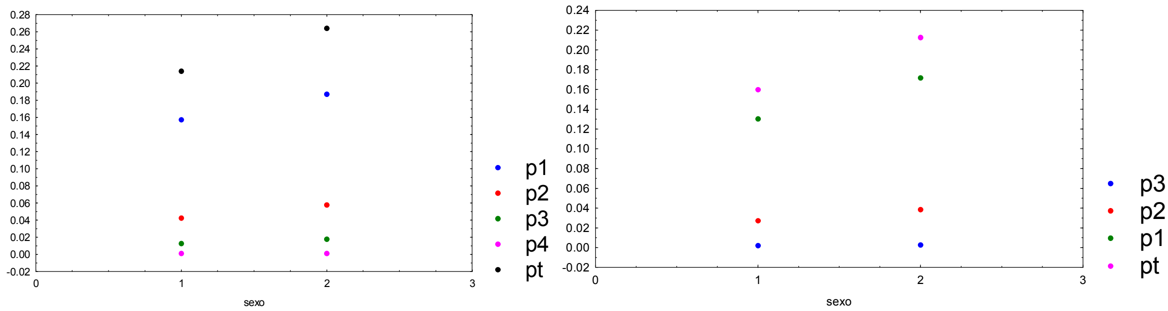


Figure 2. Graphics of probability of occurrence: a) number of deficiencies, b) visual disability according to their severity degrees for the sex variable.

The graphics of Figure 2 in all cases the increased risk of incidence of a) deficiency and b) visual disability are higher for females.

Since, Figure 3 presents the incidence risk a) number of deficiencies and b) visual disability due to age.

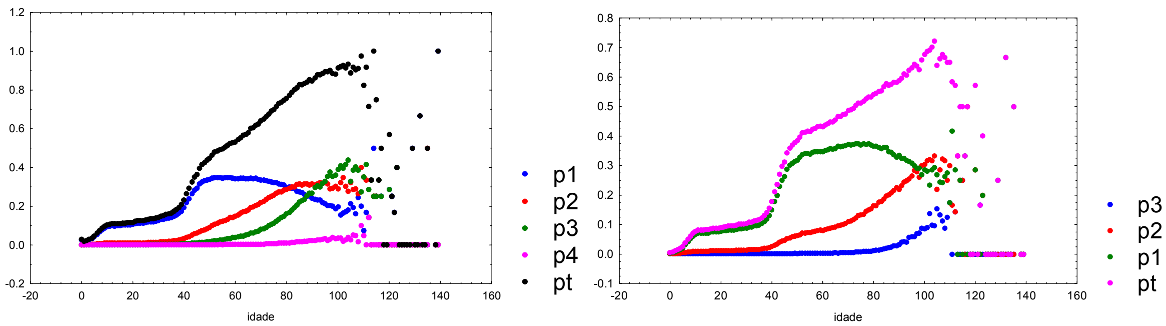


Figure 3. Graphics of the probability of occurrence of a) number of deficiencies, b) the occurrence of visual disability according to their severity degrees for age variable.

In Figure 3 you can see that the risks in the a) disability and b) visual disability that increase as the age of the people interviewed.

Note also that, after a certain age start to randomize the points, this type of occurrence. It is believed that it is due to a smaller number of persons interviewed at higher ages and this occurs mainly above 80 years of age.

In Figure 4 we consider the races: 1 – White, 2 - black, 3 - yellow, 4 - brown and 5 - indigenous.

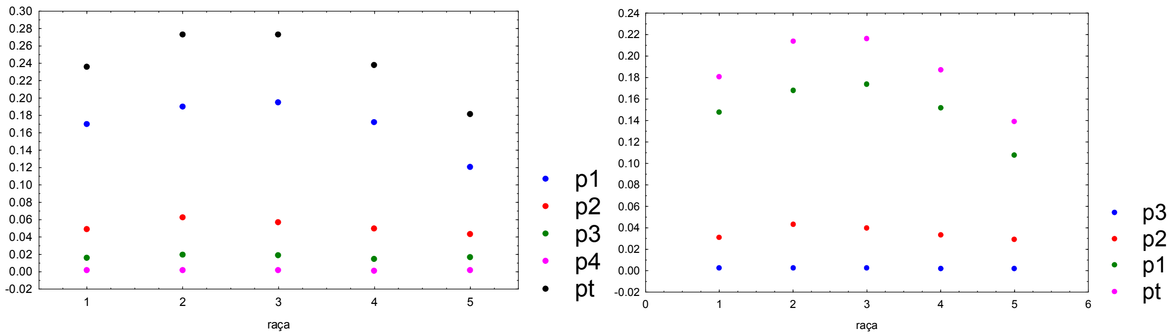


Figure 4. Graphics of the probability of occurrence of a) number of deficiencies, b) visual disabilities according to their severity degrees for variable race.

The results of Figure 4 we see that higher probabilities of incidence of disabilities and visual disabilities is in the yellow race and lower incidence in indigenous race.

In the Figure 5, we consider the following levels of education 1 - "*between uneducated and incomplete elementary school*", 2 - "*full primary and incomplete secondary level*", 3 - "*between complete higher school and incomplete college*" and finally 4 - "*full college level or more*."

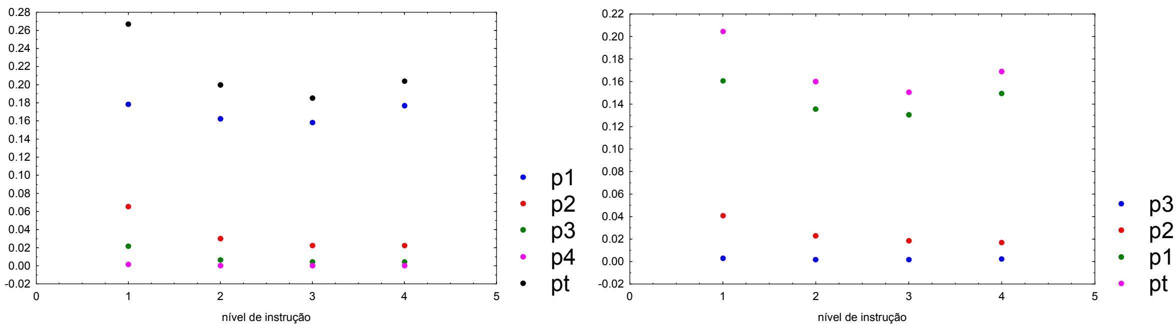
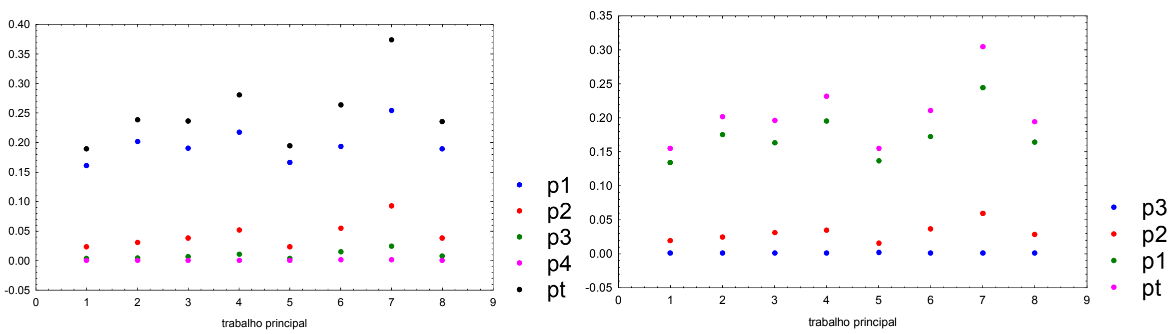


Figure 5. Graphics of the probability of occurrence of a) number of deficiencies, b) visual disabilities according to their severity degrees for education level.

In Figure 5 we observe that the higher incidence of risk and disability risk visual disability is found in 1, "*among uneducated and incomplete elementary school*", while the lower incidence of these risks is found in 3 "*full of middle and upper level incomplete*" in all situations.

For Figure 6 we consider the following levels for major work: 1 - "*employees with a formal contract*", 2 - "*Military and civil servants*", 3 - "*workers without a formal contract*", 4 - "*own*", 5 - "*employers*", 6 - "*unpaid*", 7 - "*workers producing for their own consumption*", and, finally, 8 - "*total*".



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 February 2018 doi:10.20944/preprints201802.0171.v1

Figure 6. Graphics of the probability of occurrence of a) number of deficiencies, b) visual disability according to their severity degrees for major work.

Looking at the graphics in Figure 6, we found that the greatest risk of incidence of disabilities and visual disability are at 7, "*workers producing for their own consumption*" and lower risk of incidence in both cases were found in 2 "*employees with a formal contract*."

For the income situation in Figure 7 established 1 - "*0 to 1 minimum wage*", 2 - "*between 1 and 3 minimum wages*", 3 - "*between 3 and 7 minimum wages*", 4- "*between 7 and 15 minimum wages*", and finally, 5 - "*15 minimum wages or more*¨.

By the results in the graphics in Figure 7, we see that the greatest risk of incidence of disabilities and visual disability was found in one "*between 0 and 1 minimum wage*" and note that this risk decreases to the extent that income increases the interviewee.

Finally, in Figure 8 was made the scatter plot for the risk of incidence of disabilities and visual disability in the number of children.

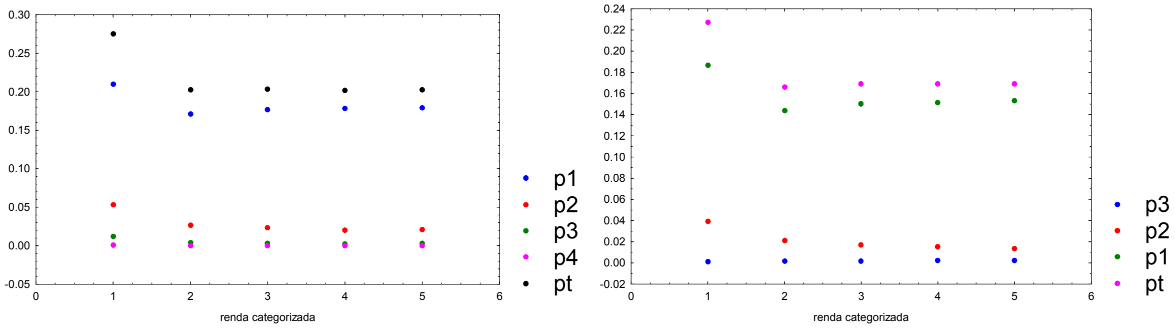


Figure 7. Graphics of the probability of occurrence of a) number of deficiencies, b visual disabilities according to their severity degrees categorized for income.

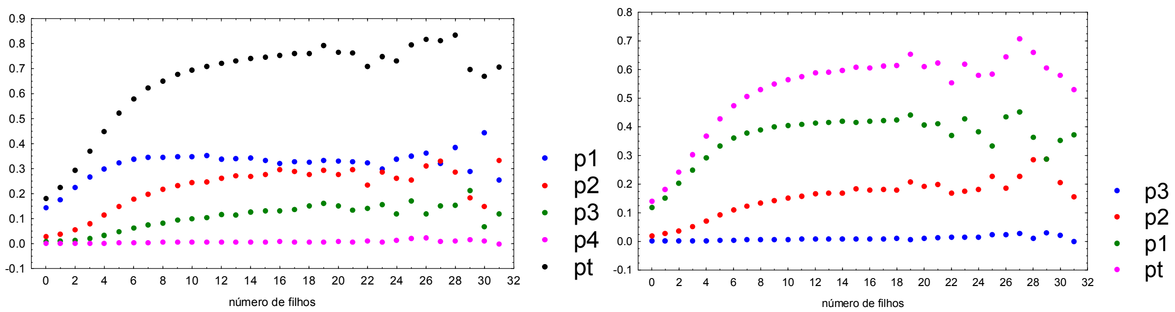


Figure 8. Graphics of the probability of occurrence of a) number of deficiencies, b) visual disability occurrence according to their severity degree for many children.

In the case of Figure 8, it is generally observed that the risk of incidence of disabilities and visual disability increases as the number of children increases.

Tables 1-5 show the results of the stereotype ordinal logistic regression analyzes; models selection by AIC, BIC and DIC criteria, and, point and interval estimates of the parameters considering as response.

The number of different disability (Table 1), visual disability (Table 2), hearing (Table 3), walking (Table 4) and intellectual (Table 5) which are all marked in bold, and the explanatory variables included in the final model for each of the different considered significant deficiencies in accordance with the method of stepwise - backward.

For variable number of deficiencies, we obtain the following significant predictors as fit for each of the different blocks:

- Identification: household, categorized age, birthplace, nationality and region;
- Education: reading and writing, childcare, other graduate and level education;
- Family: marriage nature, marital status and number of children categorized;
- Labor: income, derivative work, main work, time of job and return; and, finally;

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 February 2018 doi:10.20944/preprints201802.0171.v1

- End model: (Table 1) (formed from by the connection of all predictor variables that were significant in each of the previous blocks): region, naturally, read and write, daycare, occupation situation, education level, marriage nature, marital status, number of children categorized, income, return and main work.

For model selection, we obtain AIC = -7232.953, BIC = -8791.418 and DIC = -6917.953.

For visual disability, were selected for each block, following variables:

Identity: Region, home, sex, place of birth and nationality;

- Education: reading and writing, childcare, other graduate and education level;
- Family: marriage nature, marital status and number of children categorized;
- Labor: income, length, condition, main and secondary work; and, finally;
- End model: (Table 2) initialized with all explanatory variables that were significant for each of the different blocks and have been selected: region, naturally, reading and writing, childcare, education level, marriage nature, number of children categorized, return, condition and situation.

For model selection, we obtain: AIC = -2549.708, BIC = -3291.833 and DIC = -2399,707.

For hearing disability, were selected for each block, following variables:

- Education: read and write, daycare, other graduate and instruction level;
- Family: marriage nature, civil status and number of children;
- Labor: income, length, condition, location, primary and secondary work;
- Identification: region, domicile, gender, race and naturalness; and, finally;
- For model selection (Table 3): region, naturally, reading and writing, instruction level, marital status, children, condition and situation.

For model selection we obtain: AIC = -2921.348, BIC = -3331.401 and DIC = -2865,348.

For move disability:

- Identification: region, age categorized, and naturalness;
- Education: reading and writing, childcare, other graduate and education level;
- Family: marriage nature, civil status and children;
- Labor: income, return, time, condition, location, primary and secondary work, and finally;
- End model (Table 4): region, naturally, reading and writing, childcare, instruction level, marital status, children, return, time, condition and main job.

For model selection we obtain: AIC = -1258.613, BIC = -2119.480 and DIC = -1084,013.

Finally, in the case of Intellectual disability, it was selected as significant variables for each block:

- Identification: region, gender, age categorized, race and place of birth;
- Education: reading and writing, childcare, other graduate and education level;
- Family: nature of marriage, marital status, and children;
- Labor: income, length, return, condition, location and secondary job; and, finally;
- Final model (Table 5): sex, age categorized, naturally, reading and writing, and instruction level.

To obtain model selection: AIC = -14 548, -14 711 = BIC and DIC = -14 515.

Making a comparative study between the models contained in Tables 1-5 we note that the model that included fewer variables was adjusted logistic model for intellectual disability, whereas the model that required the largest number of independent variables was for number of deficiencies.

 doi:10.20944/preprints201802.0171.v1

Table 1. Point and interval parameters estimates of the stereotype ordinal logistic model considering as dependent variable the number of deficiencies (deficiencies).

variáveis		Estimativa	Erro padrão	Wald	df	Sig.	Intervalo de confiança 95%	
							Limite inferior	Limite superior
	0	**-.210**	**.075**	**7.786**	**1**	**.005**	**-.358**	**-.063**
deficiências	1	**1.922**	**.075**	**650.417**	**1**	**.000**	**1.774**	**2.070**
	2	**3.938**	**.077**	**2636.890**	**1**	**0.000**	**3.787**	**4.088**
	3	**7.034**	**.104**	**4562.783**	**1**	**0.000**	**6.830**	**7.238**
	1	.250	.013	365.872	1	.000	.225	.276
região	2	.295	.011	714.216	1	.000	.273	.317
	3	-.071	.010	51.179	1	.000	-.090	-.051
	4	-.181	.011	281.285	1	.000	-.202	-.160
	5	0			0			
	1	-.060	.006	97.127	1	.000	-.072	-.048
naturalidade	2	.076	.014	31.279	1	.000	.049	.103
	3	0			0			
	1	-.428	.015	869.893	1	.000	-.456	-.399
ler e escrever	2	0			0			
	1	-.022	.023	.953	1	.329	-.066	.022
creche	2	-.114	.024	21.574	1	.000	-.162	-.066
	3	-.012	.019	.412	1	.521	-.050	.025
	4	0			0			
	1	.086	.059	2.130	1	.144	-.029	.201
	2	-.205	.059	12.091	1	.001	-.320	-.089
condição	3	-.405	.059	47.319	1	.000	-.520	-.289
de ocupação	4	-.486	.059	67.149	1	.000	-.602	-.370
	5	0[a]			0			
	1	.037	.014	7.143	1	.008	010	.063
nível de	2	-.060	.014	17.007	1	.000	-.088	-.031
instrução	3	-.006	.017	.118	1	.731	-.039	.028
	4	0			0			
	1	.297	.014	473.400	1	.000	.270	.324
natureza	2	.480	.024	411.506	1	.000	.434	.527
da união	3	.561	.016	1209.135	1	.000	.529	.592
	4	.819	.023	1244.424	1	.000	.773	.864
	5	0			0			
	1	-1.055	.015	5009.257	1	0.000	-1.084	-1.026
estado	2	-.909	.013	5024.643	1	0.000	-.934	-.884
civil	3	-.509	.013	1621.905	1	0.000	-.534	-.485
	4	0			0			
	1	.304	.032	89.934	1	.000	.241	.367
	2	.171	,032	29.101	1	.000	.109	.233
renda	3	.170	,032	28.047	1	.000	.107	.233
	4	.133	.035	14.378	1	.000	.064	.202

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 February 2018 doi:10.20944/preprints201802.0171.v1

665 Table 2. Point and interval parameters estimates of the stereotype ordinal logistic model considering
666 as dependent variable visual disability

Variables		estimatives	standard errors	Wald	df	p-value	Confidence Interval 95% lower limit	Confidence Interval 95% upper limit
visual disability	**1**	**-5.190**	**.072**	**5147.165**	**1**	**0.000**	**-5.332**	**-5.048**
	2	**-2.012**	**.066**	**935.571**	**1**	**.000**	**-2.141**	**-1.883**
	3	**.177**	**.066**	**7.310**	**1**	**.007**	**.049**	**.306**
	4	**10.966**	**,140**	**6174.537**	**1**	**0.000**	**10.692**	**11.239**
Region	**1**	-.272	.013	421.476	1	.000	-.298	-.246
	2	-.262	.011	544.896	1	.000	-.284	-.240
	3	.141	.010	193.114	1	.000	.121	.161
	4	.275	.011	609.290	1	.000	.253	.297
	5	0			0			
naturalness	**1**	.051	.006	67.199	1	.000	.039	.064
	2	-.077	.014	30.191	1	.000	-.105	-.050
	3	0			0			
read and write	**1**	.369	.014	710.297	1	.000	.342	.396
	2	0			0			
childcare	**1**	-.041	.022	3.940	1	.049	-.084	.002
	2	.043	.024	3.098	1	.078	-.005	.091
	3	-.018	.019	.931	1	.335	-.054	.018
	4	0			0			
instruction level	**1**	-.062	.060	1.064	1	.302	-.181	.056
	2	.226	.060	14.015	1	.000	.108	.345
	3	.425	.060	49.529	1	.000	.307	.544
	4	.460	.061	57.577	1	.000	.341	.579
	5	0			0			
union nature	**1**	-.223	.007	1108.625	1	.000	-.236	-.210
	2	-.135	.008	260.276	1	.000	-.151	-.118
	3	-.067	.016	17.005	1	.000	-.099	-.035
	4	0			0			
children	**1**	1.101	.015	5656.802	1	0.000	1.072	1.130
	2	.922	.012	5599.102	1	0.000	.898	.946
	3	.498	.012	1696.917	1	0.000	.474	.522
	4	0			0			
return	**1**	.213	.016	166.427	1	.000	.180	.245
	2	0			0			
condition	**1**	0			0			
situation	**1**	0			0			

667

 doi:10.20944/preprints201802.0171.v1

Table 3. Point and interval parameters estimates of the stereotype ordinal logistic model considering as dependent variable the response hearing disability.

Variables		estimatives	standard errors	Wald	df	p-value	Confidence Interval 95%	
							lower limit	upper limit
hearing disability	**1**	**-6.251**	**.081**	**5885.079**	**1**	**0.000**	**-6.410**	**-6.091**
	2	**-4.350**	**.079**	**3058.665**	**1**	**0.000**	**-4.504**	**-4.196**
	3	**-2.495**	**.078**	**1017.060**	**1**	**.000**	**-2.648**	**-2.342**
	4	**12.767**	**.288**	**1962.330**	**1**	**0.000**	**12.202**	**13.332**
region	**1**	-.098	.018	29.749	1	.000	-.133	-.063
	2	-.299	.015	401.518	1	.000	-.329	-.270
	3	-.015	.014	1.204	1	.273	-.043	.012
	4	.005	.015	.131	1	.718	-.024	.035
	5	0			0			
naturalness	**1**	.039	.008	22.219	1	.000	.023	.055
	2	-.110	.018	38.663	1	.000	-.145	-.075
	3	0			0			
read and write	**1**	.449	.012	1302.286	1	.000	.424	.473
	2	0			0			
instruction level	**1**	-.445	.075	35.178	1	.000	-.592	-.298
	2	-.119	.075	2.475	1	.116	-.266	.029
	3	.123	.075	2.688	1	.101	-.024	.271
	4	.292	.076	14.760	1	.000	.143	.441
	5	0			0			
marital status	**1**	-.122	.009	177.545	1	.000	-.139	-.104
	2	-.395	.021	352.262	1	.000	-.436	-.354
	3	-.468	.016	824.993	1	.000	-.500	-.436
	4	-.816	.014	3433.926	1	0.000	-.843	-.789
	5	0			0			
children	**1**	.826	.016	2645.864	1	0.000	.795	.858
	2	.709	.013	2870.222	1	0.000	.683	.735
	3	.432	.012	1207.655	1	.000	.407	.456
	4	0			0			
condition	**1**	.042	.014	9.356	1	.002	.015	.068
	2	0			0			
situation	**1**	0			0			
	2	0			0			

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 February 2018 doi:10.20944/preprints201802.0171.v1

Table 4. Point and interval parameters estimates of the stereotype ordinal logistic model considering as dependent variable to walk

Variables		estimatives	standard errors	Wald	df	p-value	Confidence Interval 95%	
							lower limit	upper limit
walk disability	1	-5.591	.125	1.987.299	1	0.000	-5.837	-5.345
	2	-2.726	.120	512.061	1	.000	-2.962	-2.490
	3	-1.146	.120	90.832	1	.000	-1.382	-.911
	4	13.027	.234	3.110.980	1	0.000	12.569	13.485
region	1	-.182	.022	71.172	1	.000	-.224	-.139
	2	-.370	.018	436.802	1	.000	-.404	-.335
	3	-.086	.016	28.375	1	.000	-.118	-.055
	4	-.047	.018	7.112	1	.008	-.082	-.013
	5	0			0			
naturalness	1	.012	.010	1.576	1	.209	-.007	.031
	2	-.157	.020	60.058	1	.000	-.197	-.117
	3	0			0			
read and write	1	.577	.017	1.171.179	1	.000	.544	.610
	2	0			0			
childcare	1	.333	.029	133.621	1	.000	.277	.390
	2	.586	.039	220.779	1	.000	.509	.664
	3	-.025	.022	1.311	1	.252	-.068	.018
	4	0			0			
instruction level	1	-.569	.109	27.157	1	.000	-.783	-.355
	2	-.104	.109	.908	1	.341	-.319	.110
	3	.267	0,109	5.963	1	.015	.053	.482
	4	.538	0,11	23.881	1	.000	.322	.754
	5	0			0			
marital status	1	-.125	.010	142.810	1	.000	-.146	-.105
	2	-.530	.022	585.604	1	.000	-.572	-.487
	3	-.621	.017	1.326.137	1	.000	-.654	-0,588
	4	-.933	.016	3.244.635	1	0.000	-.965	-.901
	5	0			0			
children	1	1.009	.020	2.554.328	1	0.000	.970	1.049
	2	.794	.016	2.471.085	1	0.000	.763	.825
	3	.382	.015	639.566	1	.000	.352	.411
	4	0			0			
return	1	0			0			
time	1	.691	.031	506.529	1	.000	.631	.751
	2	.673	.029	524.780	1	.000	.615	0,73
	3	.508	.030	283.211	1	.000	.449	.568
	4	.284	.032	80.361	1	.000	.222	.346
	5	0			0			
condition	1	0			0			
situation	1	0			0			
main job	1	.542	.033	266.774	1	.000	.477	.607
	2	.497	.241	4.229	1	.040	.023	.970
	3	-.010	.036	.076	1	.783	-.080	.060
	4	.415	.033	158.503	1	.000	.351	.480
	5	.216	.034	40.930	1	.000	.150	.283
	6	.623	.053	139.555	1	.000	.520	.727
	7	0			0			

 doi:10.20944/preprints201802.0171.v1

Table 5. Point and interval parameters estimates of the stereotype ordinal logistic model considering as dependent variable intellectual disability

Variables		estimatives	standard errors	Wald	df	p-value	Confidence Interval 95% lower limit	Confidence Interval 95% upper limit
intellectual disability	**1**	**-3.795**	**.087**	**1923.902**	**1**	**0.000**	**-3.964**	**-3.625**
	2	**10.498**	**.096**	**11945.597**	**1**	**0.000**	**10.310**	**10.686**
sex	1	-.103	.006	284.656	1	.000	-.115	-.091
	2	0			0			
age	1	.664	.013	2719.203	1	0.000	.639	.689
	2	.073	.008	84.007	1	.000	.057	.089
	3	0			0			
naturalness	1	-.139	.007	454.576	1	.000	-.152	-.126
	2	-.251	.015	274.087	1	.000	-.281	-.221
	3	0			0			
read and write	1	1.486	.007	46829.311	1	0.000	1.473	1.500
	2	0			0			
instruction level	1	-.943	.086	120.203	1	.000	-1.112	-.775
	2	-.303	.087	12.254	1	.000	-.473	-.133
	3	.116	.087	1.802	1	.179	-.054	.286
	4	.362	.089	16.647	1	.000	.188	.535
	5	0			0			

4. Discussions

4.1. Justifications for the statistical techniques, procedures and databases used in this research.

Database usage of the demographic type is justified because it contains the set of data and variables that best represents the country. Allowing in cases like this one to investigate the different deficiencies together of other variables collected such as level of education, income, marital status, type of work among others, obtaining more comprehensive results representing all the country.

In this article, we propose the creation of this index, which will be composed of the weighting of the responses of the different variables obtained from the microdata of the 2010 IBGE Census, since it is the most comprehensive database with regard to information on the Brazilian population. This methodology gradually emerges from the simplest statistical techniques more elaborate such as multivariate techniques such as the logistic regression model.

The epidemiological view in this article is related to social aspects, accessibility and assistive technology among others, and physicians, under the policy of prevention, treatment and health control.

The need to measure well-being appeared in the mid-twentieth century as consequence of growing concern about the quality of life that resulted from the consequences of the industrialization process. However, as research has progressed in this area, new reasons have arisen that justify the need to assess the well-being or quality of life, among which consider the creation of a risk index for persons with disabilities (risk of any disabled person to better assess the quality of life of these people.

Quality of life indicators are condensed, simplified, and qualified information that facilitates communication, comparisons and the decision-making process. Quality of life indicators are also proposed as an incentive for the mobilization of society to pressure those who make decisions.

A question of interest today is to evaluate the quality of life, whether of an individual, municipality, state, region or country. One of the points that formed the evaluation of this life quality formed by the set of variables as education level, transportation time, income, age, sex and others. These variables may be depending on their value, increase or decrease the risk disability that can serve as an evaluation tool and in the implementation of public policies directed preventively at the whole population. To direct the work so that the group of people with disabilities is better included in society and its different duties and rights.

Given this scenario, reasons have arisen that justify the need to evaluate the well-being or life quality of the disabilities people. We propose the creation of the risk index person with disability composed by weighing the responses of the different variables obtained from the Census IBGE and selected as significant after applying the stepwise backward methodology in polygonal logistic regression adjustment for each deficiency studied, and; [25] and binary logistic regression [1]

In the case of a person becoming disabled, it is related to the future after a certain time interval so that the people involved can have better management, security, social promotion, health promotion and accessibility.

The use of variable selection and model selection techniques allows a better understanding of the variables that influence the different deficiencies and multiple deficiencies and which model best fits, which can help in the implementation of public policies that can prevent or inhibit several factors that increase the number of cases of different disabilities, which contributes to these people gaining more in terms of inclusion and accessibility.

The creation of the disabled person risk index made available, monitored in real time and repeated on a continuous basis can facilitate the work to be done by the public authorities offering solutions faster and efficiently.

4.2. About the results obtained in this research.

The following Figures 1-8 to pass on to establish possible reasons and suggestions for papers or research that can accept or reject the hypotheses.

• Figure 1, these results can be justified by the low effective investment in health and infrastructure that are lower in northeastern regions and higher in other regions such as southeast and south.

So you can better evaluate this hypothesis, can make a survey of the effective volume required in each of these areas in health, accessibility and infrastructure that benefit disabilities people among the different regions accounting for the amount of people that were actually processed and do a study comparison among the different regions;

• Figure 2: Most likely, these results reflect increased exposure of women to domestic accidents and the double shift of the modern woman in work out and take care of the family.

To better appreciate a point a study sample of working time at home and away from home for men and women can be done and make a comparative study;

• Figure 3: These results show that with the passage of time, the population is more aged with longer life expectancy and thus more subject to diseases of old age and a higher incidence of being disabled.

In this case, we suggest that prove the one hand the increase of life expectancy of the population, and on the other hand, occurrence of diseases that tend to occur at older ages and this can be done by considering the data of Sample IBGE Census 2010;

• Figure 4: These results may reflect the cultural and dietary conditions of Eastern and indigenous peoples.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 February 2018 doi:10.20944/preprints201802.0171.v1

For a better understanding of this result, we suggest a research study on the habits of life of individuals from the yellow and indigenous breeds, considering the chances of becoming disabled or not;

• Figure 5: If you believe that, these results may be due to factors such as: low education level can mean less knowledge of information, low purchasing power and greater dependence on government aid.

In this case, to establish, encourage research that can establish relationships between education level and income that most likely will get this kind of result;

• Figure 6: Most likely, the different types of functions reflect the education level obtained by different workers, be they military or statutory require approval in a public contest that requires a higher education level, and while working on own consumption in general, consists of people who work in the field and with lower purchasing power.

To better understand and evaluate this hypothesis suggest that a survey can establish the average pay for each of the different professions.

• Figure 7: The justification is because the incidence disability risk or visual disability risk and lower the higher is the purchasing power of people.

In this case, we suggest to do a study on a survey of disabilities people and who don´t have disabilities and then, visual disability and who don´t have visual disability and do a comparison among different income levels, and finally;

• Figure 8: this result may reflect situations as more children can mean a higher number of accidents and less parental attention for each child in social and economic terms.

This result may reflect situations as more children can mean a higher number of accidents and less parental attention for each child in social and economic terms.

In this case, what might be interesting would be to establish a research to compare the quality of life among families with different numbers of children and make comparisons among different cases in terms of disability risk.

The results in Figures 1-8 were similar for number of disabilities and visual disability.

The results of this study also show the need for further studies, research and analysis, for when it comes to risk; we note that there are numerous methods to determine this risk, either using the regression coefficients, the disability risk score factor considering the risk weighting for each of the different explanatory variables. For example, it is known that the risk deficiency increases as age increases, so does the number of children, and so on.

4.3. Future studies

Among other alternatives that are worth being considered for future work, include:

1 - To consider the beta regression, factor analysis, structural equation modeling and the BART algorithm that in all these cases offer opportunities for interesting work as well as ways to improve the fit quality and its reliability in the determination of this index style.

2 - Repeat analysis including related to housing conditions and possession of other property like car, TV and computer with internet variables.

3 - Among the various issues that are required to be answered is questions about how disabilities people live and in what situation they find themselves to be compared to people without disabilities.

4 - In situations like is a risk index with good reliability and fit quality is of interest to facilitate the monitoring of the situation, just as occurs with the HDI (Human Development Index), although the latter is an index more generally, not yet consider the issue of disabilities people.

5 - Another topic of interest would be to evaluate the accessibility of the villas for disabilities persons, given the locations of these residence georeferenced form evaluating the conditions of the surrounding infrastructure, for this, we propose a geostatistical modeling that can incorporate issues like this.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 February 2018 doi:10.20944/preprints201802.0171.v1

6 - The problem that exists when solves calculate a risk index is to consider a method that is efficient and reliable and to decrease the risk of discrepancy, for reasons such as these; it is of interest that several methods should be considered.

7 - The advantage exists an index like this is that it can be comparable, i.e., it can be evaluated and be studied to understand why his value has increased or decreased. Such that, the higher the index, should reflect a greater need for intervention by the government so you can lower the barriers that exist in terms of access to different human and affordability of housing near the disabled rights.

8 - Propose improvements in the IBGE census questionnaire, for example, if a respondent answered that a disabled person, is also interesting that he answer at what age it was acquired, since according to the literature [2], it is known that, the older he become disability person, biggest are the difficulties that this person for adapt.

9 - In statistical terms:

- we need to improve national statistics on disability, employing an efficient and cost effective to obtain more comprehensive data approach and add questions on disability;
- Statistical connection among different data sets, collecting longitudinal data, including questions on disability, so you can make better monitoring, improve the data comparability;
- Developing appropriate tools and fill the gaps between research and, ultimately;
- Strengthen and support the different investigations considering creating instruments that can measure and monitor the life quality and well-being of these people continuously and periodically.

10 - Consider also include issues related to health, accessibility and leisure.

11 - Repeat the analysis by region, state and municipalities.

12 - Results of studies like this can assist in the action of public managers offering better support the care of disabilities persons.

13 - Propose to the IBGE, improvements to be made to census data regarding the best use and possibilities of other types of analysis that concerns the conditions of people with disabilities such as: ask if a respondent has a disability with which age was acquired and if they respond age zero to interpret that it was acquired at birth and its respective CIF. These additions are motivated that according to experts when the disability was acquired at birth has better possibilities of adaptation from when it is later acquired and the ICF to standardize the national data according to international protocol established by WHO and protocol of Washington.

14 - Critical analysis of the manual WHO/ESCAP Training Manual on Disability Statistics.

5. Conclusions

The model required a smaller number of explanatory variables was mental or intellectual 5, whereas what was required larger amount to 13 for response variable disabilities.

In this work, by using stereotype ordinal logistic model with ordinal response, it was possible to improve the fit quality, to be compared with the adjustment proposed by [1] with binary response logistic model and multinomial or polyatomic. By using ordinal response merged the disability risk for different severity degree and amount of deficiencies.

Different disabilities aren´t homogenous as the different predictor variables.

The incidence risks of being disabled person and more specially a visual disability person are probably higher in situations such as: living in the northeast, female, age 80, yellow race, instruction level until fundamental incomplete, works production for their own consumption and has a high number of children.

The lower incidence risks are noted in situations with living in Southern, male, age maximum 15 years indigenous race, instruction level between complete fundamental and incomplete mid-level worker on the payroll and has no children.

Acknowledgments: The author thanks the IBGE for access to microdata for the selected composition of the sample and by Professor Maria Pavan Soler indication of the subject themes.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 February 2018 doi:10.20944/preprints201802.0171.v1

Referencies

1. OLIVEIRA, P.T.M.S. Pessoas com deficiência: análise dos resultados do Censo 2010 e a sua evolução. In: *58 RBRAS/ 15 SEAGRO*, in the period between July 22 to 26 2013, Campina Grande – PB, 2013.

2. GARCIA, V.G. *Pessoas com deficiência e o mercado de trabalho*, Doutoral Thesis. Economy Institut – UNICAMP, Campinas – SP, 2010.

3. FIGUEIRA, E. *Caminhando em silêncio*, Giz editorial and Livraria Ltda., São Paulo-SP, 2008.

4. FILHO, N.A.; BARRETO, M.L. *Epidemiologia & Saúde. Fundamentos, métodos e aplicações*. Guanabara Koogan, 2012, Rio de Janeiro-RJ.

5. BITTENCOURT, H.R. Regressão logística politômica: revisão teórica e aplicações. *Aeta Scientiae*, **2003**, 5(1):77-86.

6. OLIVEIRA, P. T. M. S. Disabilities people: risk issue from polytomous logistic regression application and under epidemiological view. In: XIV Regression Models School, from March 2 to 5, Convention Center, UNICAMP, Campinas-SP, 2015.

7. Anderson J.A. Regression and ordered categorical variables. *J R Statisti Soc*, **1984**,16: 1-30.

8. ABREU, M.N.S. *Uso de modelos de regressão logística ordinal em epidemiologia: um exemplo usando a qualidade de vida*. Master Thesis. School of Public Health, UFMG, Belo Horizonte - MG, 2007.

9. HASTIE, T.; TIBSHIRANI, P.; FRIEDMANN, J. *The elements of learning: data mining, inference and population*. Springer, canada, 2009.

10. OLIVEIRA, P.T.M.S. *Aplicação do algoritmo genético no mapeamento de genes epistáticos em cruzamentos controlados*. Doctoral Thesis, IME-USP, São Paulo – SP, 2008.

11. CASELLA, G.; BERKER, P.L. *Statistical inference*, Wadsmorth, Brooks, California, EUA, 1990.

12. OLIVEIRA, P.T.M.S. *Estimação e testes de hipóteses em calibração comparativa*, Master Thesis, IME-USP, São Paulo – SP, 2001.

13. DRAPER, N.R.; SMITH, H. *Applied regression analysis*, John Wiley, New York, 1998.

14. CAMARINHA-FILHO, J.A. *Modelos lineares mistos: estimativas de matrizes de variância e covariâncias e seleção de modelos,* Doctoral Thesis, ESDALQ-USP, Piracicaba-SP, 2008.

15. BROMAN, K, W. *Identifying Quantitative Trait Locos in Experimental Cross*, Thesis, University of California, Berkeley, 1997.

16. BURNHAM, K.P.; ANDERSON, D.R. *Model selection and inference*, Springer, New York, USA, 1998.

17. BURNHAM, K.P.; ANDERSON, D.R. *Model selection and multimodel inference*, Springer, New York, USA, 2002.

18. PAULINO, C.D., TURKMAN, A.A., MURTEIRA, B.J.F. *Estatística Bayesiana*, Fundação Calouste Gulbenkian, Portugal, 2003.

 doi:10.20944/preprints201802.0171.v1

19. SAKAMOTO, Y.; ISHGURU, M.; KITAMURA, G. *Akaike information criterion statistics*, KTK, Scientific Publisher, Japan, 1986.

20. SUSER, M. (1987). Epidemiology, Health & Society – selected papers. New York: Oxford University press.

21. ALMEIDA FILHO, N.; ROUQUAYROL, M.Z. (2006) Introdução a epidemiologia moderna. Guanabara Koogan, Rio de Janeiro-RJ.

22. SPINK, M.J.; ARAGAKI, S.S.; ALVES, M.P. Da exacerbação dos sentidos no encontro com a natureza: contratando esportes radicais e turismo de aventura. *Psicologia Relex Critical*. **2005**, 18:26-38.

23. LUIZ, O.C.; COHN, A, *Sociedade de risco e risco epidemiológico*, Cadernos Saúde Pública, Rio de Janeiro-RJ, 2006.

24. SKOLBEKKE, J. The risk epidemic in medical journals. *Soc Sci Med*.**1995**, 40:291-405.

25. OLIVEIRA, P.T.M.S. (2014). Ser pessoa com deficiência: questão de risco. *Revista da Estatística UFOP*, **2014**, 3(2):228 – 253.