RT-MAE 9620

ASYMPTOTIC PROPERTIES OF THE MSAE ESTIMATORS IN THE DOSE-RESPONSE MODEL

by

Carmen D.S. André, Subhash C. Narula, Clovis A. Peres and Gilberto A. Ventura

Palavras-Chave: minimum sum of absolute errors regression, Monte Carlo,
(Key words) multistage model, nonlinear model.

Classificação AMS: 62E20, 62E25.
(AMS Classification)

- Setembro de 1996 -

ASYMPTOTIC PROPERTIES OF THE MSAE ESTIMATORS IN THE DOSE-RESPONSE MODEL

Carmen D. S. André
University of Sao Paulo, Sao Paulo, Brazil

Subhash C. Narula
Virginia Commonwealth University, Richmond, Virginia, USA

Clovis A. Peres
Paulista School of Medicine, Sao Paulo, Brazil

Gilberto A. Ventura
University of Sao Paulo, Sao Paulo, Brazil

Abstract

The multistage dose response model is frequently used in practical applications. The minimum sum of absolute errors MSAE criterion is more resistant to outliers than the popular least squares procedure. Efficient algorithms have been proposed to compute the MSAE estimates of the parameters of the model. However, the properties of these estimators are not known. In this paper, our objective is to study asymptotic properties of these estimators. We also give an approximate expression for the variance of these estimators when their asymptotic distribution is multinormal.

Key words: minimum sum of absolute errors regression, Monte Carlo, multistage model, nonlinear model.

1. INTRODUCTION

For the carcinogenesis risk assessment studies, the multistage model proposed by Armitage and Doll (1954)

$$P(d) = 1 - \exp(-\sum_{j=0}^{k} \alpha_j d^j), \qquad (1)$$

is used most frequently, where k is the number of stages and $\alpha_j \geq 0$ are the parameters, $j = 0, 1, \dots, k$. Peres (1981) derived the model for radiobiological data where the response variable was the survival to a dose d of radiation. The probability at dose zero is called background response, and the highest dose which will not produce acute toxicity is called the maximum tolerated dose, MTD. The maximum likelihood and the least squares methods are often used to estimate the parameters of (1); however, these may be unduly affected by outliers, André, Peres, and Narula (1991). They proposed the minimum sum of absolute errors MSAE estimators for the parameters of (1) because these estimators are more resistant to outliers than the least squares estimators.

Let $(d_1, y_1), \ldots, (d_n, y_n)$ represent the values of dose-response in a bioassay with N experimental units at each of the n doses. Then (1) may be written as:

$$y_i = 1 - \exp(-\sum_{j=0}^{k} \alpha_j d_i^j) + e_i, \qquad (2)$$

where e_i is the unobservable random error for the i^{th} observation. In order to minimize the sum of absolute errors for (2), we minimize

$$\sum_{i=1}^{n} |e_i| = \sum_{i=1}^{n} |y_i - 1 + \exp(-\sum_{j=0}^{k} \alpha_j d_i^j)|. \qquad (3)$$

Because the function is intrinsically nonlinear, it may be solved by using any algorithm for computing the MSAE estimates of a nonlinear model, Gonin and Money (1987). André, Peres and Narula (1991,1992) proposed two algorithms which exploit the special structure of the model to compute the estimates. However, at present, the statistical properties of these estimators are not known.

In this paper, our objective is to investigate the asymptotic properties of these estimators as N approaches infinity. The rest of the paper is organized as follows: in Section 2, we describe the Monte Carlo study and in Section 3, we present and discuss the results. In Section 4, we give an expression for the asymptotic variance of the estimators. We conclude the paper with few observations and remarks in Section 5.

2. THE MONTE CARLO STUDY

Guess, Crump and Peto (1977) and Land (1980) reported that in practical applications to chemical carcinogens and risk analyses for ionizing radiation, the multistage models have at most two stages. In this paper, we limit our study to only the following one stage model:

$$P(d) = 1 - \exp(-\alpha_0 - \alpha_1 d). \qquad (4)$$

Setting MTD equal to one, we obseve that the parameters α_0 and α_1 have a direct and clear interpretation in this model. For example, α_0 is approximately equal to P(0) for small values of α_0 and α_1 is proportional to P(1), see (5a) and (5b),

$$\alpha_0 \cong P(0) = 1 - \exp(-\alpha_0) \qquad (5a)$$

and

$$\alpha_1 = -\alpha_0 - \ln(1 - P(1)). \qquad (5b)$$

Let y_i denote the value of the response corresponding to d_i, the i^{th} dose, $i = 1, 2, \ldots, n$, where n denotes the number of doses. For the observed values, (4) may be written as:

$$y_i = 1 - \exp(-\alpha_0 - \alpha_1 d_i) + e_i, \qquad (6)$$

where e_i denotes the residual for the i^{th} observation. We want to determine the asymptotic distribution of the MSAE estimators of the parameters in (6). To do so, we conducted a Monte Carlo study for fixed values of α_0 and α_1. We needed to fix these values, or equivalently, the values of P(0) and P(1), see Table 1, such that the results of the study will be widely applicable.

Table 1. (P(0), P(1)) and corresponding values of (α_0, α_1)

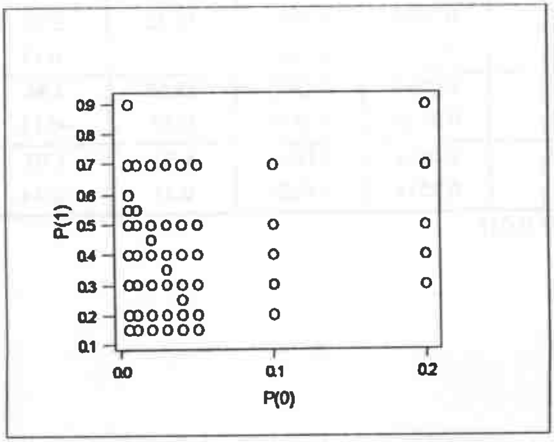
P(0)	P(1)	α_0	α_1
0.005	0.30	0.005	0.3517
0.005	0.90	0.005	2.2976
0.100	0.30	0.105	0.2567
0.100	0.90	0.105	2.2026

Based on the historical dose-response observed by the National Cancer Institute's bioassay program, Portier (1982) concluded that in practical problems the values of P(0) and P(1) lie in the following intervals:

$$0.001 \leq P(0) \leq 0.30 \qquad \text{and} \qquad 0.15 \leq P(1) \leq 0.90.$$

The values of P(0) and P(1) used in the study are displayed in Figure 1. We fixed the number of doses n equal to five at the following values: $d_1 = 0.00$, $d_2 = 0.25$, $d_3 = 0.50$, $d_4 = 0.75$, and $d_5 = 1.00$. At each dose d_i, the response follows the binomial distribution with parameters N and P(d_i) where N is the number of experimental units observed at each dose. To determine the asymptotic properties and distributions of the MSAE estimators, we considered various values of N between 50 and 10 000. For each model and N, we generated and estimated 1500 dose-response curves using the MSAE criterion. For the Monte Carlo distributions of $\hat{a}_0$ and $\hat{a}_1$, we computed the mean, the standard deviation, the kurtosis, the skewness and tested their normality using the Kolmogrov-Smirnov test statistic and critical values of Lilliefors (1967).

Figure 1. Values of P(0) and P(1) used in the simulation study



3. RESULTS AND DISCUSSION

The results for selected values of the parameters and sample sizes are displayed in Table 2. Similar results were observed for other values of the parameters and sample sizes.

Table 2a. Means, standard deviations, kurtosis and skewness of the distributions of $\hat{\alpha}_0$ and $\hat{\alpha}_1$ generated from the true curve : $P(d) = 1 - \exp(-0.0050 - 2.2976d)$, and observed values of the Kolmogorov-Smirnov statistic.

N	Estimator	Mean	Std Dev	Kurtosis	Skewness	K-S(*)
100	$\hat{\alpha}_0$	0.0050	0.0070	2.71	1.49	0.3262
	$\hat{\alpha}_1$	2.3407	0.1949	0.47	0.50	0.0511
200	$\hat{\alpha}_0$	0.0052	0.0050	1.21	1.02	0.2154
	$\hat{\alpha}_1$	2.3313	0.1330	0.19	0.26	0.0298
500	$\hat{\alpha}_0$	0.0050	0.0032	0.97	0.75	0.1623
	$\hat{\alpha}_1$	2.3114	0.0830	0.07	0.28	0.0242
1000	$\hat{\alpha}_0$	0.0050	0.0023	0.06	0.42	0.1052
	$\hat{\alpha}_1$	2.3075	0.0605	0.10	0.36	0.0300
10000	$\hat{\alpha}_0$	0.0050	0.0007	0.11	0.05	0.0214
	$\hat{\alpha}_1$	2.2979	0.0194	0.01	0.12	0.0216

(*) 5% critical value = 0.031

Table 2b. Means, standard deviations, kurtosis and skewness of the distributions of $\hat{\alpha}_0$ and $\hat{\alpha}_1$ generated from the true curve : $P(d) = 1 - \exp(-0.0050 - 0.3517d)$, and observed values of the Kolmogorov-Smirnov statistic.

N	Estimator	Mean	Std Dev	Kurtosis	Skewness	K-S(*)
100	$\hat{\alpha}_0$	0.0079	0.0123	15.35	3.15	0.2599
	$\hat{\alpha}_1$	0.3517	0.0529	0.90	0.05	0.0253
200	$\hat{\alpha}_0$	0.0067	0.0086	9.41	2.58	0.2170
	$\hat{\alpha}_1$	0.3521	0.0370	0.16	-0.10	0.0202
500	$\hat{\alpha}_0$	0.0060	0.0062	13.32	2.98	0.1648
	$\hat{\alpha}_1$	0.3515	0.0245	0.15	-0.14	0.0173
1000	$\hat{\alpha}_0$	0.0055	0.0042	15.68	2.84	0.1357
	$\hat{\alpha}_1$	0.3512	0.0170	0.12	-0.15	0.0113
10000	$\hat{\alpha}_0$	0.0051	0.0016	6.95	1.03	0.1168
	$\hat{\alpha}_1$	0.3515	0.0024	0.35	-0.14	0.0158

(*) 5% critical value = 0.031

Table 2c. Means, standard deviations, kurtosis and skewness of the distributions of $\hat{\alpha}_0$ and $\hat{\alpha}_1$ generated from the true curve : P(d) = 1 - exp(-0.2231 - 2.0794d), and observed values of the Kolmogorov-Smirnov statistic.

N	Estimator	Mean	Std Dev	Kurtosis	Skewness	K-S(*)
100	$\hat{\alpha}_0$	0.2238	0.0503	-0.02	0.30	0.0595
	$\hat{\alpha}_1$	2.1540	0.2290	1.46	0.59	0.0480
200	$\hat{\alpha}_0$	0.2224	0.0352	0.29	0.30	0.0505
	$\hat{\alpha}_1$	2.1262	0.1532	0.42	0.33	0.0344
500	$\hat{\alpha}_0$	0.2232	0.0230	-0.01	0.15	0.0382
	$\hat{\alpha}_1$	2.1050	0.1022	0.17	0.25	0.0252
1000	$\hat{\alpha}_0$	0.2232	0.0165	0.04	0.14	0.0344
	$\hat{\alpha}_1$	2.0934	0.0707	0.31	0.41	0.0367
10000	$\hat{\alpha}_0$	0.2232	0.0050	-0.12	0.04	0.0190
	$\hat{\alpha}_1$	2.0810	0.0221	-0.02	0.11	0.0186

(*) 5% critical value = 0.031

Table 2d. Means, standard deviations, kurtosis and skewness of the distributions of $\hat{\alpha}_0$ and $\hat{\alpha}_1$ generated from the true curve : P(d) = 1 - exp(-0.2231 - 0.1335d), and observed values of the Kolmogorov-Smirnov statistic.

N	Estimator	Mean	Std Dev	Kurtosis	Skewness	K-S(*)
100	$\hat{\alpha}_0$	0.1966	0.0348	-0.19	0.11	0.0450
	$\hat{\alpha}_1$	0.1986	0.0581	0.52	0.51	0.0499
200	$\hat{\alpha}_0$	0.2139	0.0280	0.11	-0.02	0.0269
	$\hat{\alpha}_1$	0.1587	0.0467	-0.09	0.41	0.0573
500	$\hat{\alpha}_0$	0.2227	0.0206	0.03	0.12	0.0213
	$\hat{\alpha}_1$	0.1365	0.0361	-0.19	0.10	0.0241
1000	$\hat{\alpha}_0$	0.2237	0.0148	0.20	0.06	0.0189
	$\hat{\alpha}_1$	0.1335	0.0262	0.35	0.09	0.0211
10000	$\hat{\alpha}_0$	0.2233	0.0045	0.30	0.02	0.0197
	$\hat{\alpha}_1$	0.1338	0.0079	0.11	0.03	0.0179

(*) 5% critical value = 0.031

It may be observed that the distribution of $\hat{\alpha}_1$ always approaches normal. The distribution of $\hat{\alpha}_0$ approaches normal distribution when α_0 is not very close to the boundary ($\alpha_0 > 0.03$) for all values of α_1. For small values of α_0 ($\alpha_0 < 0.02$), the distribution of $\hat{\alpha}_0$ approaches the normal distribution only when the response at dose one (P(1)) is high.

Based on the results of the Monte Carlo study, we may conclude the following:

1. The MSAE estimators $\hat{\alpha}_0$ and $\hat{\alpha}_1$ are asymptotically unbiased.

2. The MSAE estimator $\hat{\alpha}_1$ is asymptotically normally distributed.

3. The region, over which the asymptotic distribution of $\hat{\alpha}_0$ is normal, is displayed in Figure 2. Note that for the sake convenience, the region is displayed in the (P(0), P(1)) space.

4. When the asymptotic distribution of $\hat{\alpha}_0$ is not normal, it is approximately symmetric about the mean and its kurtosis is greater than that of normal distribution as displayed in Figure3.

Figure 2- Region where the distribution of $\hat{\alpha}_0$ is normal

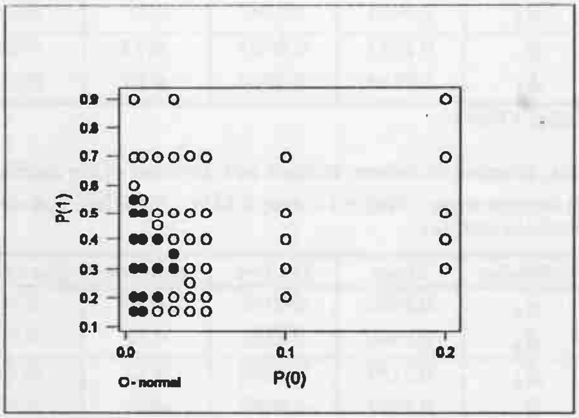
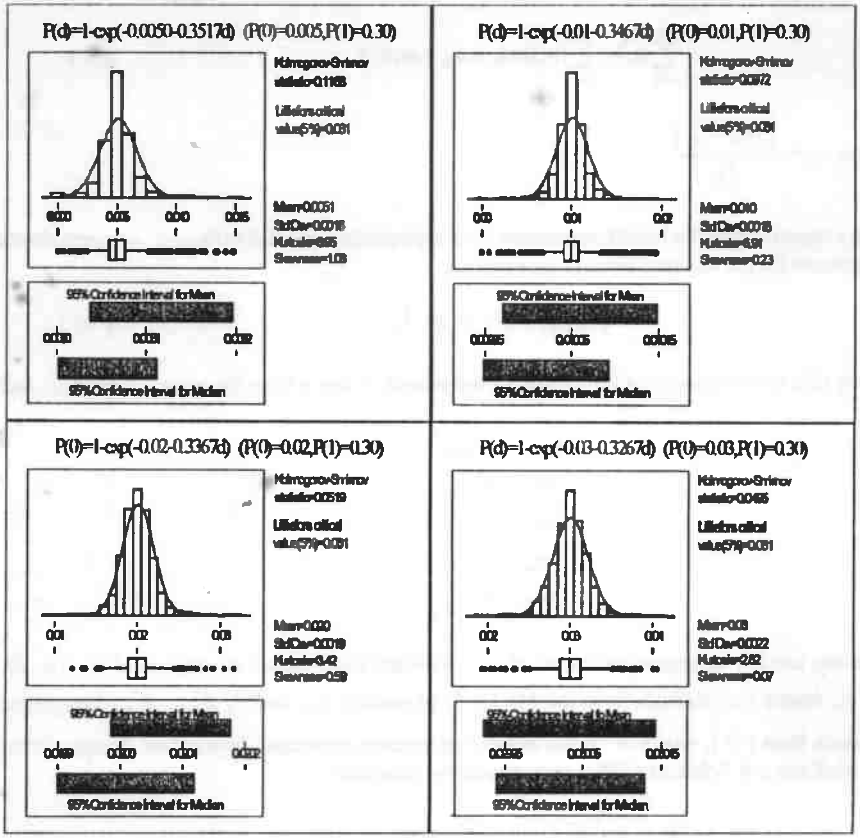


Figure 3. Asymptotic distribution of $\hat{\alpha}_0$ for some values of α_0 and α_1 when it is not normal



4. ASYMPTOTIC VARIANCE

André (1981) has shown that model (1) can be approximated by a linear model and the errors from (6) may be written as:

$$|\, e_i \,| \cong |\, z_i(lnz_i + \alpha_0 + \alpha_1 d_i)| \qquad (7)$$

where zi = 1 - yi, i = 1, 2, ..., n. Let

$$\frac{|e_i|\sqrt{N}}{\sqrt{y_i(1-y_i)}} \cong \frac{\sqrt{N}}{\sqrt{y_i(1-y_i)}}\,|\, z_i(lnz_i + \alpha_0 + \alpha_1 d_i)| = |e_i^*|. \qquad (8)$$

Observe that e_i^*'s are the errors of a linear model.

Now, we minimize:

$$\sum_{i=1}^{n} |e_i^*| = \sum_{i=1}^{n} |w_i(\ln z_i + \alpha_0 + \alpha_1 d_i)|, \tag{9}$$

where $w_i = \dfrac{\sqrt{N(1-y_i)}}{\sqrt{y_i}}$.

In the region where the MSAE estimators have asymptotic normal distribution, an approximate expression for the variance-covariance matrix is:

$$Var(\hat{\alpha}) = \tau^2 (X'X)^{-1}, \tag{10}$$

where τ2/n is the variance of the median of the sample of size n from the error distribution, and

$$X = \begin{bmatrix} w_1 & w_1 d_1 \\ w_2 & w_2 d_2 \\ w_3 & w_3 d_3 \\ w_4 & w_4 d_4 \\ w_5 & w_5 d_5 \end{bmatrix}.$$

One may use any consistent estimator of τ to estimate the variance covariance of $\hat{\alpha}$ Let, $\hat{e}_1$, $\hat{e}_2, \ldots, \hat{e}_n$ denote the residuals from the MSAE fit of model (6) and $\hat{e}_1^*, \hat{e}_2^*, \ldots, \hat{e}_{n^*}^*$ denote the residuals from (9), where n^* is the number of nonzero residuals. Birkes and Dodge (1993) and McKean and Schrader (1984) recommend the estimator:

$$\hat{\tau} = \sqrt{n^*}(\hat{e}^*_{(n^*-m+1)} - \hat{e}^*_{(m)})/4.$$

where $m = (n^* + 1)/2 - \sqrt{n^*}$.

5. SUMMARY

The results of the Monte Carlo study show that the MSAE estimators of the single stage dose response model are asymptotically unbiased, $\hat{\alpha}_1$ follows a normal distribution, and $\hat{\alpha}_0$ follows a normal distribution over certain region of the parameter space. We have also given an approximate expression for the variance of the estimators when they follow normal distribution. Using these results it will be possible to develop some statistical inference procedures.

ACKNOWLEDGMENT

We thank FAPESP for the grant that partially supported the work of Professor Narula and CNPq for the grant that supported the work of Mr. Ventura.

REFERENCES

André, C. D. S.(1989). Estimação L_1 em um modelo de dose resposta. Ph.D. Dissertation, Institute of Mathematics and Statistics, University of Sao Paulo, Sao Paulo, Brazil.

André, C. D. S. ; Peres, C. A. and Narula, S. C.(1991). An iterative procedure for the estimation of parameters in a dose-response model. *Communications in Statistics, 20*, 736-776.

André, C. D. S. ; Peres, C. A. and Narula, S. C.(1992). An algorithm for the estimation of the multistage dose-response model. *Computational Statistics and Data Analysis, 14*,293-299.

Armitage, P. and Doll, R.(1954). The age distribution of cancer and a multistage theory of carcinogenesis. *British Journal of Cancer, 8*,1-12

Birkes, D. and Dodge, Y. (1993). *Alternative Methods of Regression*. John Wiley and Sons, Inc., New York.

Gonin, R. and Money, A. H.(1987). A review of computational methods for solving the nonlinear L_1-norm estimation problem. *Statistical Data Analysis on the L_1-norm and related methods*(In: Y. Dodge, ed.). Elsevier Science Publishers, Netherlands.

Guess, H. ; Crump, K. and Peto, R. (1977). Carcinogenic risk assessment: a guide to literature. *Biometrics, 37*, 353-366.

Land, C. (1980). Estimating cancer risks from low dose of ionizing radiation. *Science, 209*, 1197-1203.

Lilliefors, H. W. (1967). On the Kolmogrov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association, 62, 339-402.*

McKean, J. W. and Schrader, R. M. (1984). A comparison of methods for studentizing the sample median. *Communications in Statistics - Simulation and Computations, B13*, 751-773.

Peres, C. A. (1981). Derivação de um modelo para a análise de efeitos biológicos induzidos por radiação. Tese para concurso de livre-docência, Institute of Mathematics and Statistics, University of Sao Paulo, Sao Paulo, Brazil.

Portier, C.(1981) .Optimal bioassay design under the Armitage-Doll multistage model, Ph.D Dissertation, University of North Carolina, Chapel Hill, North Carolina, USA.

ÚLTIMOS RELATÓRIOS TÉCNICOS PUBLICADOS

9601 - MENTZ, R.P.; MORETTIN, P.A. and TOLOI, C.M.C. Bias correction for estimators of the residual variance in the ARMA (1,1) Model. São Paulo, IME-USP, 1996. 21p. (RT-MAE-9601)

9602 - SINGER, J.M.; PERES, C.A.; HARLE, C.E. Performance of Wald's test for the Hardy-Weiberg equilibrium with fixed sample sizes. São Paulo, IME-USP, 1996.15p. (RT-MAE-9602).

9603 - SINGER, J.M. and E. SUYAMA, E. Dispersion structure, Hierarchical models, Random effects models, Repeated measures. São Paulo, IME-USP, 1996. 21p. (RT-MAE-9603).

9604 - LIMA, A.C.P. and SEN, P.K. A Matrix-Valued Counting Process with First-Order Interactive Intensities. São Paulo, IME-USP, 1996, 25p. (RT-MAE-9604)

9605 - BOTTER, D.A. and SINGER, J.M. Experimentos com Intercâmbio de Dois Tratamentos e Dois Períodos: Estratégias para Análise e Aspectos Computacionais, São Paulo, IME-USP, 1996. 18p. (RT-MAE-9605)

9606- MORETTIN, P.A. From Fourier to Wavelet Analysis of Time Series. São Paulo, IME-USP, 1996. 11p. (RT-MAE-9606)

9607 - SINGER, J.M.; HO, L.L. Regression Models for Bivariate Counts. São Paulo, IME-USP, 1996. 20p. (RT-MAE-9607)

9608 - CORDEIRO, G.M. and FERRARI, S.L.P. A method of mements for finding Bartlett-type corrections. São Paulo, IME-USP, 1996. 11p. (RT-MAE-9608)

9609 - VILCA-LABRA, F.; ARELLANO-VALLE, R.B.; BOLFARINE, H. Elliptical functional models. São Paulo, IME-USP, 1996. 20p. (RT-MAE-9609)

9610 - BELITSKY, V.; FERRARI, P.A.; KONNO, N. A Refinement of Harris-FKG Inequality for Oriented Percolation. São Paulo, IME-USP, 1996. 13p. (RT-MAE-9610)

9611 - GIMENEZ, P.; BOLFARINE, H. Unbiased Score Functions in Error-In-Variables Models. São Paulo, IME-USP, 1996. 23p. (RT-MAE-9611)

9612 - BUENO, V.C. Comparing component redundancy allocation in K-out-of-n system. IME-USP, 1996. 10p. (RT-MAE-9612)

9613 - GALEA, M.; PAULA, G.A.; BOLFARINE, H. Local influence in elliptical linear regression models. IME-USP, 1996. 14p. (RT-MAE-9613)

9614 - LIMA, C.R.; BOLFARINE, H.; SANDOVAL, M.C. Linear calibration in multiplicative measurement error models. IME-USP, 1996. 12p. (RT-MAE-9614)

9615 - AVERBACH, M.; CUTAIT, R.; WECHSLER, S.; CORRÊA, P.; BORGES, J.L.A. Aplicação da inferência bayesiana no estudo das probabilidades de diagnóstico, por colonoscopia, das afecções colorretais em portadores da síndrome da imunodeficiência adquirida com diarréia. IME-USP, 1996. 10p. (RT-MAE-9615)

9616 - YOSHIDA, O.S.; LEITE, J.G.; BOLFARINE, H. Inferência bayesiana do número de espécies de uma população. IME-USP, 1996. 33p. (RT-MAE-9616)

9617 - CARMONA, S.C.; TANAKA, N.I. Exponential estimates for "Not Very Large Deviations" and ware front propagation for a class of reaction-diffusion equations. IME-USP, 1996. 30p. (RT-MAE-9617)

9618 - IRONY, T.Z.; PEREIRA, C.A.B.; TIWARI, R.C. On The Comparison Between Two Correlated Proportions in 2 x 2 Tables. IME-USP, 1996. 15p. (RT-MAE-9618)

9619 - ANDRÉ, C.D.S.; ELIAN, S.N.; NARULA, S.C.; AUBIN, E.C. Stepwise procedures for selecting variables in the minimum sum of absolute errors regression. IME-USP, 1996.9p. (RT-MAE-9619)

The complete list of "Relatórios do Departamento de Estatística", IME-USP, will be sent upon request.

Departamento de Estatística
IME-USP
Caixa Postal 66.281
05389-970 - São Paulo, Brasil